

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 16.4783 Seconds
(without alignments)
1190.526 Million cell updates/sec

Title: US-09-804-690-4

Perfect score: 2002

Sequence: 1 MVSYKYRDLTVRETNVIT.....FQLRALMQKARKTAGLSLDLY 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents, AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2002	100.0	380	1	US-08-585-758A-4
2	2002	100.0	380	1	US-08-977-818-4
3	2002	100.0	380	2	US-08-670-274B-4
4	2002	100.0	380	3	US-09-146-187-4
5	2002	100.0	390	2	US-08-786-999-1
6	2002	100.0	390	4	US-09-216-387-1
7	2002	100.0	390	4	US-09-886-319A-2
8	1900.5	94.9	381	1	US-08-585-758A-2
9	1900.5	94.9	381	1	US-08-977-818-2
10	1900.5	94.9	381	2	US-08-670-274B-2
11	1900.5	94.9	381	2	US-08-786-999-3
12	1900.5	94.9	381	3	US-09-146-187-2
13	1900.5	94.9	381	4	US-09-216-387-3
14	1900.5	94.9	391	4	US-09-886-319A-1
15	1404	70.1	285	4	US-09-886-319A-82
16	1117	55.8	237	3	US-08-999-774A-8
17	160	8.0	466	3	US-08-526-136-13
18	148.5	7.4	148	4	US-09-461-325-453
19	148.5	7.4	148	4	US-10-012-542-453
20	141	7.0	905	2	US-08-574-959A-9
21	141	7.0	905	3	US-09-357-014-9
22	141	7.0	1135	2	US-08-574-959A-7
23	141	7.0	1135	3	US-09-357-014-7
24	130	6.5	543	4	US-09-535-008-63
25	130	6.5	577	4	US-09-535-008-61
26	130	6.5	1846	4	US-09-535-008-67
27	130	6.5	1847	4	US-09-535-008-2
					Sequence 4, Appli
					Sequence 4, Appli
					Sequence 4, Appli
					Sequence 1, Appli
					Sequence 2, Appli
					Sequence 2, Appli
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 1, Appli
					Sequence 8, Appli
					Sequence 8, Appli
					Sequence 13, Appli
					Sequence 453, App
					Sequence 453, App
					Sequence 9, Appli
					Sequence 9, Appli
					Sequence 7, Appli
					Sequence 63, Appli
					Sequence 61, Appli
					Sequence 67, Appli
					Sequence 2, Appli

28	130	6.5	1649	4	US-09-535-008-75	Sequence 75, Appli
29	130	6.5	1650	4	US-09-535-008-71	Sequence 71, Appli
30	130	6.5	1678	4	US-09-535-008-69	Sequence 69, Appli
31	130	6.5	1679	4	US-09-535-008-65	Sequence 65, Appli
32	130	6.5	1681	4	US-09-535-008-77	Sequence 77, Appli
33	130	6.5	1682	4	US-09-535-008-73	Sequence 73, Appli
34	127	6.3	1184	4	US-09-266-225D-18	Sequence 18, Appli
35	127	6.3	1185	3	US-09-041-886-23	Sequence 23, Appli
36	126	6.3	1172	4	US-09-800-729-138	Sequence 138, App
37	126	6.3	868	4	US-09-800-729-106	Sequence 106, App
38	126	6.3	921	4	US-09-800-729-199	Sequence 199, App
39	126	6.3	8991	4	US-08-714-741-32	Sequence 32, Appli
40	125.5	6.3	330	1	US-08-642-255-32	Sequence 32, Appli
41	125.5	6.3	408	1	US-07-609-716-65	Sequence 65, Appli
42	125.5	6.3	408	3	US-08-475-411A-65	Sequence 65, Appli
43	125.5	6.3	408	3	US-08-478-029A-65	Sequence 65, Appli
44	124.5	6.2	1162	2	US-08-728-323A-2	Sequence 2, Appli
45	124.5	6.2	1162	4	US-09-298-568-2	Sequence 2, Appli
46	124.5	6.2	1162	4	US-09-410-399-2	Sequence 2, Appli
47	124.5	6.2	1274	4	US-09-095-443-2	Sequence 2, Appli
48	123.5	6.2	380	2	US-09-026-587-4	Sequence 4, Appli
49	123.5	6.2	380	2	US-09-227-420-4	Sequence 4, Appli
50	123.5	6.2	380	4	US-09-387-811-4	Sequence 4, Appli
51	123.5	6.2	635	4	US-09-417-197-125	Sequence 125, App
52	123	6.1	1248	2	US-09-080-897-2	Sequence 2, Appli
53	123	6.1	1248	3	US-09-323-735-2	Sequence 2, Appli
54	122	6.1	1315	3	US-08-899-595-3	Sequence 3, Appli
55	119.5	6.0	503	3	US-08-526-136-2	Sequence 2, Appli
56	119.5	6.0	505	3	US-08-526-136-4	Sequence 4, Appli
57	118.5	5.9	1255	2	US-09-080-897-4	Sequence 4, Appli
58	118.5	5.9	1255	3	US-08-899-595-1	Sequence 1, Appli
59	118.5	5.9	1255	3	US-09-323-735-4	Sequence 4, Appli
60	118	5.9	214	1	US-08-217-327-4	Sequence 4, Appli
61	117	5.8	416	4	US-09-690-454-136	Sequence 136, App
62	117	5.8	779	4	US-10-164-595-56	Sequence 56, Appli
63	117	5.8	843	4	US-10-164-595-54	Sequence 54, Appli
64	116.5	5.8	338	1	US-08-218-686-2	Sequence 2, Appli
65	116.5	5.8	338	3	US-08-460-242-2	Sequence 2, Appli
66	116	5.8	1461	4	US-09-585-887-9	Sequence 9, Appli
67	116	5.8	1461	4	US-09-289-578-9	Sequence 9, Appli
68	116	5.8	1464	4	US-09-331-347C-21	Sequence 21, Appli
69	115.5	5.8	357	1	US-07-609-716-66	Sequence 66, Appli
70	115.5	5.8	357	1	US-08-642-255-33	Sequence 33, Appli
71	115.5	5.8	357	3	US-08-475-411A-66	Sequence 66, Appli
72	115.5	5.8	357	3	US-08-478-029A-66	Sequence 66, Appli
73	115	5.7	559	4	US-10-116-370-2	Sequence 2, Appli
74	114.5	5.7	1017	4	US-09-600-776-6	Sequence 6, Appli
75	113.5	5.7	97	4	US-09-599-360B-118	Sequence 118, App
76	113.5	5.7	561	1	US-08-642-255-52	Sequence 52, Appli
77	113.5	5.7	777	1	US-08-642-255-53	Sequence 53, Appli
78	113.5	5.7	1418	3	US-08-963-825-20	Sequence 20, Appli
79	113.5	5.7	1418	4	US-09-500-811-20	Sequence 20, Appli
80	113.5	5.7	1418	4	US-09-570-573-20	Sequence 20, Appli
81	113.5	5.7	1418	4	US-09-548-608-20	Sequence 20, Appli
82	113	5.6	668	4	US-09-277-431A-2	Sequence 2, Appli
83	113	5.6	902	1	US-08-396-479B-6	Sequence 6, Appli
84	113	5.6	902	1	US-08-818-823-6	Sequence 6, Appli
85	112	5.6	2441	1	US-08-194-468-2	Sequence 2, Appli
86	112	5.6	2441	3	US-08-961-739-2	Sequence 2, Appli
87	112	5.6	2441	4	US-09-514-247A-8	Sequence 8, Appli
88	112	5.6	2441	4	US-09-686-316-2	Sequence 2, Appli
89	111.5	5.6	723	6	5200183-4	Patent No. 5200183
90	111.5	5.6	809	5	PCT-US91-01726-3	Sequence 3, Appli
91	111.5	5.6	1341	3	US-08-963-825-18	Sequence 18, Appli
92	111.5	5.6	1341	4	US-09-500-811-18	Sequence 18, Appli
93	111.5	5.6	1341	4	US-09-570-573-18	Sequence 18, Appli
94	111.5	5.6	1341	4	US-09-548-608-18	Sequence 18, Appli
95	111.5	5.6	1418	3	US-09-010-999-1	Sequence 1, Appli
96	111	5.5	297	3	US-08-580-545B-6	Sequence 6, Appli
97	111	5.5	531	3	US-09-262-653A-6	Sequence 6, Appli
98	110.5	5.5	297	4	US-09-134-001C-4920	Sequence 4920, Ap
99	110	5.5	262	4	US-08-311-731A-333	Sequence 333, App
100	110	5.5	652	6	5202236-13	Patent No. 5202236

ALIGNMENTS

RESULT 1
US-08-585-758A-4
; Sequence 4, Application US/08585758A
; Patent No. 5679523
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N.
; TITLE OF INVENTION: METHOD FOR CONCURRENT DISRUPTION OF
; TITLE OF INVENTION: EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: FOUR EMBARCADERO CENTER, SUITE 3400
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/585,758A
; FILING DATE: 12-JAN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION/DOCKET NUMBER: A62783/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-585-758A-4

Query Match 100.0%; Score 2002; DB 1; Length 380;
Best Local Similarity 100.0%; Pred. No. 6.1e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSKEYRDLTVRETNVITLYKDLKPLVDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
Db 1 MVSKEYRDLTVRETNVITLYKDLKPLVDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
QY 61 PICLLDTPYPPNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHWHKHPQSDLLGLIQV 120
Db 61 PICLLDTPYPPNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHWHKHPQSDLLGLIQV 120
QY 121 MIVFGEDEPVFSRPIASYPYQATGPNTSYMPGMPGGISPYBSGYPNPSPGPGCPY 180
Db 121 MIVFGEDEPVFSRPIASYPYQATGPNTSYMPGMPGGISPYBSGYPNPSPGPGCPY 180
QY 181 PGCGPYPATSSQYPSQPPVTVVPSRDGTISEDTIRASLISAVSDKLRWKEEMDRQA 240
Db 181 PGCGPYPATSSQYPSQPPVTVVPSRDGTISEDTIRASLISAVSDKLRWKEEMDRQA 240
QY 241 AELNALKRTEEDLKKGHQKLEEMVTRLDQFAEVDKNIELKKKDELSALEKMNQSE 300
Db 241 AELNALKRTEEDLKKGHQKLEEMVTRLDQFAEVDKNIELKKKDELSALEKMNQSE 300
QY 301 NNDIDVVIPTAPLYKQILNLYAEENAIETIFVLGEALRRGVLDLDFLKHVRLLSRQK 360
Db 301 NNDIDVVIPTAPLYKQILNLYAEENAIETIFVLGEALRRGVLDLDFLKHVRLLSRQK 360

QY 361 FOLRALMOKARKTAGLSLDLY 380
Db 361 FOLRALMOKARKTAGLSLDLY 380

RESULT 2
US-08-977-818-4
; Sequence 4, Application US/08977818
; Patent No. 5807995
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N.
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/977,818
; FILING DATE: 25-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-977-818-4

Query Match 100.0%; Score 2002; DB 1; Length 380;
Best Local Similarity 100.0%; Pred. No. 6.1e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSKEYRDLTVRETNVITLYKDLKPLVDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
Db 1 MVSKEYRDLTVRETNVITLYKDLKPLVDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
QY 61 PICLLDTPYPPNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHWHKHPQSDLLGLIQV 120
Db 61 PICLLDTPYPPNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHWHKHPQSDLLGLIQV 120
QY 121 MIVFGEDEPVFSRPIASYPYQATGPNTSYMPGMPGGISPYBSGYPNPSPGPGCPY 180
Db 121 MIVFGEDEPVFSRPIASYPYQATGPNTSYMPGMPGGISPYBSGYPNPSPGPGCPY 180
QY 181 PGCGPYPATSSQYPSQPPVTVVPSRDGTISEDTIRASLISAVSDKLRWKEEMDRQA 240
Db 181 PGCGPYPATSSQYPSQPPVTVVPSRDGTISEDTIRASLISAVSDKLRWKEEMDRQA 240
QY 241 AELNALKRTEEDLKKGHQKLEEMVTRLDQFAEVDKNIELKKKDELSALEKMNQSE 300
Db 241 AELNALKRTEEDLKKGHQKLEEMVTRLDQFAEVDKNIELKKKDELSALEKMNQSE 300

QY 301 NNDIDEVIPTAPYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 360
DB 301 NNDIDEVIPTAPYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 360

QY 361 FQLRALMOKARKTAGLSLDLY 380
DB 361 FQLRALMOKARKTAGLSLDLY 380

RESULT 3

US-08-670-274B-4
; Sequence 4, Application US/08670274B
; Patent No. 5891668
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,274B
; FILING DATE: June 13, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-670-274B-4

Query Match 100.0%; Score 2002; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 6.1e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEYKRDLTVRETWNVITLYKDLKPVLDSYVFNDSGSRRLMNLGTITPVYRGNTYNI 60
DB 1 MVSKEYKRDLTVRETWNVITLYKDLKPVLDSYVFNDSGSRRLMNLGTITPVYRGNTYNI 60

QY 61 PICLLWLLDTYPNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHWHKHPQSDLLGLIQV 120
DB 61 PICLLWLLDTYPNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHWHKHPQSDLLGLIQV 120

QY 121 MIWVFGEPPVFSRPIASAPYPYQATGPPNTSYMPGMPGGISPYPSGYPNPSGYPGCPY 180
DB 121 MIWVFGEPPVFSRPIASAPYPYQATGPPNTSYMPGMPGGISPYPSGYPNPSGYPGCPY 180

QY 181 PFGGYPATTSQYPSQPPVTTVGPSSRDGTISEDITIRASLISAVSKLWRWKEEMDRQA 240
DB 181 PFGGYPATTSQYPSQPPVTTVGPSSRDGTISEDITIRASLISAVSKLWRWKEEMDRQA 240

QY 241 AEINALKRTEEDLKKGHOKLEEMVTRLDQVAVVDKNIELLKKKDEELSSALEKMEENOSE 300

DB 241 AEINALKRTEEDLKKGHOKLEEMVTRLDQVAVVDKNIELLKKKDEELSSALEKMEENOSE 300
QY 301 NNDIDEVIPTAPYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 360
DB 301 NNDIDEVIPTAPYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 360
QY 361 FQLRALMOKARKTAGLSLDLY 380
DB 361 FQLRALMOKARKTAGLSLDLY 380

RESULT 4

US-09-146-187-4
; Sequence 4, Application US/09146187
; Patent No. 6248523
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-146-187-4

Query Match 100.0%; Score 2002; DB 3; Length 380;
Best Local Similarity 100.0%; Pred. No. 6.1e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEYKRDLTVRETWNVITLYKDLKPVLDSYVFNDSGSRRLMNLGTITPVYRGNTYNI 60
DB 1 MVSKEYKRDLTVRETWNVITLYKDLKPVLDSYVFNDSGSRRLMNLGTITPVYRGNTYNI 60

QY 61 PICLLWLLDTYPNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHWHKHPQSDLLGLIQV 120
DB 61 PICLLWLLDTYPNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHWHKHPQSDLLGLIQV 120

QY 121 MIWVFGEPPVFSRPIASAPYPYQATGPPNTSYMPGMPGGISPYPSGYPNPSGYPGCPY 180
DB 121 MIWVFGEPPVFSRPIASAPYPYQATGPPNTSYMPGMPGGISPYPSGYPNPSGYPGCPY 180

QY 181 PFGGYPATTSQYPSQPPVTTVGPSSRDGTISEDITIRASLISAVSKLWRWKEEMDRQA 240

Db 181 PPGPYPATISSQVPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKLWRMKEEMDRQ 240
Qy 241 AELNALKRTBEDLKKGHQKLEEMVTRLDQVAVVDKNIELLLKKDEBLSALEKMNQSE 300
Db 241 AELNALKRTBEDLKKGHQKLEEMVTRLDQVAVVDKNIELLLKKDEBLSALEKMNQSE 300
Qy 301 NNDIDEVIIPTAPLYKQILNLYABENAIEDTIFVLGEALRGVIDLDVFLKHVLLSRKQ 360
Db 301 NNDIDEVIIPTAPLYKQILNLYABENAIEDTIFVLGEALRGVIDLDVFLKHVLLSRKQ 360
Qy 361 FQRLMQKARKTAGLSLDY 380
Db 361 FQRLMQKARKTAGLSLDY 380

RESULT 5

US-08-786-999-1
; Sequence 1, Application US/08786999
; Patent No. 5892016
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: Filed Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Colnot01
; CLONE: 609476

US-08-786-999-1

Query Match 100.0%; Score 2002; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 6.4e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSVKYKRLTVRETNVITLYKDLKPLVDSYVFDGSSRELMLNLTGTPVPVPGNTYNI 60
Db 11 MSVKYKRLTVRETNVITLYKDLKPLVDSYVFDGSSRELMLNLTGTPVPVPGNTYNI 70
Qy 61 PICLWLLDTPYNPFCFVKPTSMITKCKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 71 PICLWLLDTPYNPFCFVKPTSMITKCKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy 121 MIWVFGDEPPVFRSPISASYPPOATGPPNTSYMPGMPGGISPYSPGPPNPSPGCPY 180
Db 131 MIWVFGDEPPVFRSPISASYPPOATGPPNTSYMPGMPGGISPYSPGPPNPSPGCPY 190
Qy 181 PPGPYPATISSQVPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKLWRMKEEMDRQ 240
Db 191 PPGPYPATISSQVPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKLWRMKEEMDRQ 250
Qy 241 AELNALKRTBEDLKKGHQKLEEMVTRLDQVAVVDKNIELLLKKDEBLSALEKMNQSE 300
Db 251 AELNALKRTBEDLKKGHQKLEEMVTRLDQVAVVDKNIELLLKKDEBLSALEKMNQSE 310
Qy 301 NNDIDEVIIPTAPLYKQILNLYABENAIEDTIFVLGEALRGVIDLDVFLKHVLLSRKQ 360
Db 311 NNDIDEVIIPTAPLYKQILNLYABENAIEDTIFVLGEALRGVIDLDVFLKHVLLSRKQ 370
Qy 361 FQRLMQKARKTAGLSLDY 380
Db 371 FQRLMQKARKTAGLSLDY 390

RESULT 6

US-09-216-387-1
; Sequence 1, Application US/09216387
; Patent No. 6472508
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/216,387
; FILING DATE: 18-Dec-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
; LENGTH: 390 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Colnot01
; CLONE: 609476
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-216-387-1

Query Match 100.0%; Score 2002; DB 4; Length 390;
Best Local Similarity 100.0%; Pred. No. 6.4e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSXKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 60
DB 11 MVSXKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 70
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKVDANGKIYLYLHWHKHPQSDLLGLIQV 120
DB 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKVDANGKIYLYLHWHKHPQSDLLGLIQV 130
QY 121 MIVVFGDEPPVPSRPISASYPYQATGPPNTSYMPGMPGGISYPSPGPGCPY 180
DB 131 MIVVFGDEPPVPSRPISASYPYQATGPPNTSYMPGMPGGISYPSPGPGCPY 190
QY 181 PPGGPYPATTSQQSPQSPVPTTVTGGPSRDGTISEDTIRASLISAVSDKLRWKMKEMDRAQ 240
DB 191 PPGGPYPATTSQQSPQSPVPTTVTGGPSRDGTISEDTIRASLISAVSDKLRWKMKEMDRAQ 250
QY 241 AELNALKRTEDDLKKGHOKLEEMVTRLDQVAVDKNIELLLKKKDEELSSALEKMNQSE 300
DB 251 AELNALKRTEDDLKKGHOKLEEMVTRLDQVAVDKNIELLLKKKDEELSSALEKMNQSE 310
QY 301 NNDDEVIPTAPLYKQILNLYAEENATEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 360
DB 311 NNDDEVIPTAPLYKQILNLYAEENATEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 370
QY 361 FQRLALMQARKTAGLSLY 380
DB 371 FQRLALMQARKTAGLSLY 390
RESULT 7
US-09-886-319A-2
; Sequence 2, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the diagnosis or treatment of skin disorders and wound
; TITLE OF INVENTION: Healing and for the identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-319A-2

Query Match 100.0%; Score 2002; DB 4; Length 390;
Best Local Similarity 100.0%; Pred. No. 6.4e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSXKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 60
DB 11 MVSXKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 70
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKVDANGKIYLYLHWHKHPQSDLLGLIQV 120
DB 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKVDANGKIYLYLHWHKHPQSDLLGLIQV 130
QY 121 MIVVFGDEPPVPSRPISASYPYQATGPPNTSYMPGMPGGISYPSPGPGCPY 180

DB 131 MIVVFGDEPPVPSRPISASYPYQATGPPNTSYMPGMPGGISYPSPGPGCPY 190
QY 181 PPGGPYPATTSQQSPQSPVPTTVTGGPSRDGTISEDTIRASLISAVSDKLRWKMKEMDRAQ 240
DB 191 PPGGPYPATTSQQSPQSPVPTTVTGGPSRDGTISEDTIRASLISAVSDKLRWKMKEMDRAQ 250
QY 241 AELNALKRTEDDLKKGHOKLEEMVTRLDQVAVDKNIELLLKKKDEELSSALEKMNQSE 300
DB 251 AELNALKRTEDDLKKGHOKLEEMVTRLDQVAVDKNIELLLKKKDEELSSALEKMNQSE 310
QY 301 NNDDEVIPTAPLYKQILNLYAEENATEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 360
DB 311 NNDDEVIPTAPLYKQILNLYAEENATEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 370
QY 361 FQRLALMQARKTAGLSLY 380
DB 371 FQRLALMQARKTAGLSLY 390

RESULT 8
US-08-585-758A-2
; Sequence 2, Application US/08585758A
; Patent No. 5679523
; GENERAL INFORMATION:
; APPLICANT: Li, Limin
; APPLICANT: COHEN, Stanley N.
; TITLE OF INVENTION: METHOD FOR CONCURRENT DISRUPTION OF
; TITLE OF INVENTION: EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: FOUR EMBARCADERO CENTER, SUITE 3400
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,758A
; FILING DATE: 12-JAN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A62783/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-585-758A-2

Query Match 94.9%; Score 1900.5; DB 1; Length 381;
Best Local Similarity 94.5%; Pred. No. 1.1e-146;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVSXKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 60
DB 1 MVSXKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 60
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKVDANGKIYLYLHWHKHPQSDLLGLIQV 120
DB 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKVDANGKIYLYLHWHKHPQSDLLGLIQV 120
QY 121 MIVVFGDEPPVPSRPISASYPYQATGPPNTSYMPGMPGGISYPSPGPGCPY 179

Db 121 MIVFGEPPVFSRPTVSASYPVATGPPNTSYMPGMPGSIAYPSGYPNPSGYGCP 180
QY 180 YPGGPPYPATSSQYPSQPPVTVGPRDGTISDITRASLISAVSDKLWRMKEEMDRA 239
Db 181 YPPAGPYPATSSQYPSQPPVTVGPRDGTISDITRASLISAVSDKLWRMKEEMDGA 240
QY 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNOS 299
Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNOS 300
QY 300 ENNDIDEVIPTAPLYKQILNLVAEENAIEDTTFYLGEALRRGVLDLDFLKHVRLLSRK 359
Db 301 ENNDIDEVIPTAPLYKQILNLVAEENAIEDTTFYLGEALRRGVLDLDFLKHVRLLSRK 360
QY 360 QFQRLMOKARKTAGLSLDLY 380
Db 361 QFQRLMOKARKTAGLSLDLY 381

RESULT 9

US-08-977-818-2
; Sequence 2, Application US/08977818
; Patent No. 5807995
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; APPLICATION NUMBER: US/08/977,818
; FILING DATE: 25-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-977-818-2

Query Match 94.9%; Score 1900.5; DB 1; Length 381;
Best Local Similarity 94.5%; Pred. No. 1.1e-146;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MVSQKYRDLTVRETQVNTVITLYKDLKPVLDVSYVNDGSSRELNMNLTGTIPVYRGNTYNI 60
Db 1 MMSKYRDLTVRQTVNTVIAVYKDLKPVLDVSYVNDGSSRELNMNLTGTIPVYRGNTYNI 60
QY 61 PICLWLLDTPYPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHVKHPQSDLLGLIQV 120

Db 61 PICLWLLDTPYPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHVKHPQSELELIQI 120
QY 121 MIVFGEPPVFSRPTVSASYPVATGPPNTSYMPGMPGSIAYPSGYPNPSGYGCP 179
Db 121 MIVFGEPPVFSRPTVSASYPVATGPPNTSYMPGMPGSIAYPSGYPNPSGYGCP 180
QY 180 YPGGPPYPATSSQYPSQPPVTVGPRDGTISDITRASLISAVSDKLWRMKEEMDRA 239
Db 181 YPPAGPYPATSSQYPSQPPVTVGPRDGTISDITRASLISAVSDKLWRMKEEMDGA 240
QY 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNOS 299
Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNOS 300
QY 300 ENNDIDEVIPTAPLYKQILNLVAEENAIEDTTFYLGEALRRGVLDLDFLKHVRLLSRK 359
Db 301 ENNDIDEVIPTAPLYKQILNLVAEENAIEDTTFYLGEALRRGVLDLDFLKHVRLLSRK 360
QY 360 QFQRLMOKARKTAGLSLDLY 380
Db 361 QFQRLMOKARKTAGLSLDLY 381

RESULT 10

US-08-670-274B-2
; Sequence 2, Application US/08670274B
; Patent No. 5891668
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; APPLICATION NUMBER: US/08/670,274B
; FILING DATE: June 13, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-670-274B-2

Query Match 94.9%; Score 1900.5; DB 2; Length 381;
Best Local Similarity 94.5%; Pred. No. 1.1e-146;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MVSQKYRDLTVRETQVNTVITLYKDLKPVLDVSYVNDGSSRELNMNLTGTIPVYRGNTYNI 60
Db 1 MMSKYRDLTVRQTVNTVIAVYKDLKPVLDVSYVNDGSSRELNMNLTGTIPVYRGNTYNI 60

Qy	61	PICLWLDDTYPNPPI	CEVKETSSMTI	TKGKHVDANGKI	YLYLHWEHKHPQSDLLIGLQV	120
Db	61	PICLWLDDTYPNPPI	CEVKETSSMTI	TKGKHVDANGKI	YLYLHDKHPRSELLELQI	120
Qy	121	MIVFGEDEPPVFSRP	-ISASYPYQ	ATGPENNTSY	MEGPMGGISPYPSGYPPNPSGYPGCP	179
Db	121	MIVFGEDEPPVFSRPT	VSASYPY	TATGPENNTSY	MEGPMPSGISAYPSGYPPNPSGYPGCP	180
Qy	180	YPGCGYPATTSOYPS	OPPVTTVCP	SDGTISEDTIRASLI	SAYSVDKLRWEMKEMDPA	239
Db	181	YPPAGYPATTSOYPS	QPPVTTVGS	DRDGTISEDTIRASLI	SAYSVDKLRWEMKEMDGA	240
Qy	240	QAEALNLRKTEEDL	KGHKLEEMV	TRLDQV	AEVDKNIELLKKDEELSSALEKWNQS	299
Db	241	QAEALNLRKTEEDL	KGHKLEEMV	TRLDQV	AEVDKNIELLKKDEELSSALEKWNQS	300
Qy	300	ENNDIDEVII	TPALYKOILNLY	AEENAI	EDTIFYLGEALRGVTDLDVFLKHXVLLSRK	359
Db	301	ENNDIDEVII	TPALYKOILNLY	AEENAI	EDTIFYLGEALRRGVTDLDVFLKHXVLLSRK	360
Qy	360	QFQRLMOKARKTAG	LSLDLY			380
Db	361	QFQRLMOKARKTAG	LSLDLY			381

```

RESULT 11
US-08-786-999-3
; Sequence 3, Application US/08786999
; Patent No. 5892016
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; TITLE OF INVENTION: SUPPRESSOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: Filed Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1330330
US-08-786-999-3

```

Query Match 94.9%; Score 1900.5; DB 2; Length 381;

Best Local Similarity 94.5%; Pred. No. 1.le-146;		
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1		
QY	1 MVSQKYKRDLVRETVNVITLYKDKLPVLDSEYVNDGSSRELMLNLTGTIPVYRGNTYNI	60
DB	1 MMSQKYKRDLVTRQVNVIAWYKDKLPVLDSEYVNDGSSRELNVLTGTIPVYRGNIYNI	60
QY	61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGRHVDANGKIYLPYLHEWKHPDSDLGLLIQV	120
DB	61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGRHVDANGKIYLPYLHDKHPSSELLELIQI	120
QY	121 MIVFGDEPPVFSRP-ISAAYPPYQATGPPNTSYMPGMPGIGISYPGSGYPPNPSPGPGP	179
DB	121 MIVIFGEPPVFSRPTVSASYPPTYATGPPNTSYMPGMPGIGISAYPSGYPPNPSPGPGP	180
QY	180 YPPGGYPATTSQYPSQPPVTYTVGPRSDGTISEDTIRASLISAVSDKLWRMKEMDPA	239
DB	181 YPPAGYPATTSQYPSQPPVTYTVGPRSDGTISEDTIRASLISAVSDKLWRMKEMDPA	240
QY	240 QAEINALKRTEEDLKKGHOKLEEMVTRLDQVEAEVDKNIELLKKDEELSAALEKWNOS	299
DB	241 QAEINALKRTEEDLKKGHOKLEEMVTRLDQVEAEVDKNIELLKKDEELSAALEKWNOS	300
QY	300 ENNDIDEVIITPAPLYKQILNLYAEENAIEDTIFYLGEALRRGVTDLDVFLKHVLLSRK	359
DB	301 ENNDIDEVIITPAPLYKQILNLYAEENAIEDTIFYLGEALRRGVTDLDVFLKHVLLSRK	360
QY	360 QFQURALMQARKTAGLSLDLY	380
DB	361 QFQURALMQARKTAGLSLDLY	381

RESULT 12
US-09-146-187-2
; Sequence 2, Application US/09146187
; Patent No. 6248523
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```
; MOLECULE TYPE: protein
US-09-146-187-2

Query Match          94.9%; Score 1900.5; DB 3; Length 381;
Best Local Similarity 94.5%; Pred. No. 1.1e-146;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVSQYKRDLTVRQTVNVIYKDLKPVLDYVFNDSRELNNLTCTIPVYRGNTYNI 60
Db 1 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYVFNDSRELNLVLTCTIPVYRGNTYNI 60

QY 61 PICLWLLDTPYNNPICFVAPTSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 61 PICLWLLDTPYNNPICFVAPTSMTIKTKGHVDANGKIYLPYLHDWKHPRSELLELIQI 120

QY 121 MIVVFGDEPPVFSRPP-ISASYPYQATGPNTSYMPGPGGSIYPSPGYPNPGSGYPCP 179
Db 121 MIVIFGEPPVFSRPTVSASYPYATGPPNTSYMPGPGSGISAYPSGYPPNPGSGYPCP 180

QY 180 YPCGYPYATTSQYPSQPPVTVGPRDGTISEDTIRASLISAVSKLWRMKEEMDRA 239
Db 181 YPPAGPYATTSQYPSQPPVTVGPRDGTISEDTIRASLISAVSKLWRMKEEMDGA 240

QY 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNQS 299
Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNQS 300

QY 300 ENNDIDEVIITAPLYKQILNLYAEANAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 359
Db 301 ENNDIDEVIITAPLYKQILNLYAEANAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360

QY 360 QFQRLALMOKARKTAGLSLDLY 380
Db 361 QFQRLALMOKARKTAGLSLDLY 381

RESULT 13
US-09-216-387-3
; Sequence 3, Application US/09216387
; Patent No. 6472508
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; SUPPRESSOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/216.387
; FILING DATE: 18-Dec-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1330330
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-216-387-3

Query Match          94.9%; Score 1900.5; DB 4; Length 381;
Best Local Similarity 94.5%; Pred. No. 1.1e-146;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVSQYKRDLTVRQTVNVIYKDLKPVLDYVFNDSRELNNLTCTIPVYRGNTYNI 60
Db 1 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYVFNDSRELNLVLTCTIPVYRGNTYNI 60

QY 61 PICLWLLDTPYNNPICFVAPTSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 61 PICLWLLDTPYNNPICFVAPTSMTIKTKGHVDANGKIYLPYLHDWKHPRSELLELIQI 120

QY 121 MIVVFGDEPPVFSRPP-ISASYPYQATGPNTSYMPGPGGSIYPSPGYPNPGSGYPCP 179
Db 121 MIVIFGEPPVFSRPTVSASYPYATGPPNTSYMPGPGSGISAYPSGYPPNPGSGYPCP 180

QY 180 YPCGYPYATTSQYPSQPPVTVGPRDGTISEDTIRASLISAVSKLWRMKEEMDRA 239
Db 181 YPPAGPYATTSQYPSQPPVTVGPRDGTISEDTIRASLISAVSKLWRMKEEMDGA 240

QY 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNQS 299
Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNQS 300

QY 300 ENNDIDEVIITAPLYKQILNLYAEANAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 359
Db 301 ENNDIDEVIITAPLYKQILNLYAEANAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360

QY 360 QFQRLALMOKARKTAGLSLDLY 380
Db 361 QFQRLALMOKARKTAGLSLDLY 381

RESULT 14
US-09-886-319A-1
; Sequence 1, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; the Diagnosis or Treatment of Skin Disorders and Wound
; Healing and for the Identification of Pharmacologically
; Active Substances
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-886-319A-1
```

Query Match 94.9%; Score 1900.5; DB 4; Length 391;
Best Local Similarity 94.5%; Pred. No. 1.2e-146;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MYSKYKRDLTVRETNVITLYKDLKPVLDKSVFVNDGSSRELNLGTIPVYRGNTYNI 60
DB 11 MMSKYKRDLTVQTVNVJAMTKDLKPVLDKSVFVNDGSSRELNLGTIPVYRGNTYNI 70

QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGKHVDANGKIYLYPYLHEWKHPQSDLLGLIQV 120
DB 71 PICLWLLDTPYNPPICFVKPTSSMTIKTGKHVDANGKIYLYPYLHEWKHPRSELLELIQI 130

QY 121 MIVVGDDEPPVSRP-ISASYPYQATGPTNTSYMPGPGGSPYSPGPPNPSPGPGCP 179
DB 131 MIVVGEDEPPVSRP-TSASYPYATGPTNTSYMPGMPGSGISAYSPGPPNPSPGPGCP 190

QY 180 YPPGPGYPATTSQVPSQPPVTTVGPDRGTISEDTIRASLISAVSDKLRLWKMKEMDRA 239
DB 191 YPPAGPYATTSQVPSQPPVTTVGPDRGTISEDTIRASLISAVSDKLRLWKMKEMDGA 250

QY 240 QAEINALKRTEBDLKKGHQKLEEMVTRLDQEAQVVDKNIELKKKDEELSSALEKMNQS 299
DB 251 QAEINALKRTEBDLKKGHQKLEEMVTRLDQEAQVVDKNIELKKKDEELSSALEKMNQS 310

QY 300 ENNDIDEVITPAPLYKQILNLYAEENAIETDIFYLGEALRGVLDLDFLKHVRLLSRK 359
DB 311 ENNDIDEVITPAPLYKQILNLYAEENAIETDIFYLGEALRGVLDLDFLKHVRLLSRK 370

QY 360 QFQRLALMOKARKTAGLSLDLY 380
DB 371 QFQRLALMOKARKTAGLSLDLY 391

RESULT 15
US-09-886-319A-82
; Sequence 82, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-319A-82

Query Match 70.1%; Score 1404; DB 4; Length 285;
Best Local Similarity 100.0%; Pred. No. 1.9e-106;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 POSDLGLIQVLMVVFDEPPVSRPISASYPYQATGPTNTSYMPGPGGSPYSPGYP 169
DB 15 POSDLGLIQVLMVVFDEPPVSRPISASYPYQATGPTNTSYMPGPGGSPYSPGYP 74

QY 170 FNPSPGPGCPYPPGPGYPATTSQVPSQPPVTTVGPDRGTISEDTIRASLISAVSDKL 229
DB 75 FNPSPGPGCPYPPGPGYPATTSQVPSQPPVTTVGPDRGTISEDTIRASLISAVSDKL 134

QY 230 WRKMEEMDRAQELNALKRTBEDLKKGHQKLEEMVTRLDQEAQVVDKNIELKKKDEELS 289
DB 135 WRKMEEMDRAQELNALKRTBEDLKKGHQKLEEMVTRLDQEAQVVDKNIELKKKDEELS 194

QY 290 SALEKMNQSENNDIDEVITPAPLYKQILNLYAEENAIETDIFYLGEALRGVLDLDFV 349
DB 195 SALEKMNQSENNDIDEVITPAPLYKQILNLYAEENAIETDIFYLGEALRGVLDLDFV 254

QY 350 LKHVLLSRKQFQRLALMOKARKTAGLSLDLY 380
DB 255 LKHVLLSRKQFQRLALMOKARKTAGLSLDLY 285

RESULT 16
US-08-999-774A-8
; Sequence 8, Application US/08999774A
; Patent No. 6274312
; GENERAL INFORMATION:
; APPLICANT: Gish, Kurt C.
; APPLICANT: Seghezzi, Wolfgang
; APPLICANT: Shanahan, Frances
; APPLICANT: Lees, Emma M.
; APPLICANT: McManahan, Terrill K.
; TITLE OF INVENTION: Intracellular Regulatory Molecules;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,774A
; FILING DATE: 10-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,818
; FILING DATE: 11-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1200
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-999-774A-8

Query Match 55.8%; Score 1117; DB 3; Length 237;
Best Local Similarity 98.0%; Pred. No. 3.2e-83;
Matches 201; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MYSKYKRDLTVRETNVITLYKDLKPVLDKSVFVNDGSSRELNLGTIPVYRGNTYNI 60
DB 11 MYSKYKRDLTVRETNVITLYKDLKPVLDKSVFVNDGSSRELNLGTIPVYRGNTYNI 70

QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGKHVDANGKIYLYPYLHEWKHPQSDLLGLIQV 120
DB 71 PICLWLLDTPYNPPICFVKPTSSMTIKTGKHVDANGKIYLYPYLHEWKHPQSDLLGLIQV 130

QY 121 MIVVFGDEPPVFRSPISASYPYQATGPPNTSYMPGMPGGISPYPSGYPNBSGYPCPY 180
DB 131 MIVVFGDEPPVFRSPISASYPYQATGPPNTSYMPGMPGGISPYPSGYPNBSGYPCPY 190
QY 181 PPGGYPATTSOYPSQPPVTTVGP 205
DB 191 PPGGYPATTSOYPSQPPVLLVP 215

RESULT 17
US-08-526-136-13
; Sequence 13, Application US/08526136
; Patent No. 6107089
; GENERAL INFORMATION:
; APPLICANT: Towle, Christine A. et al.
; TITLE OF INVENTION: ANNEXIN XI
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,136
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,036
; FILING DATE:
; APPLICATION NUMBER: 07/837,775
; FILING DATE: February 13, 1992
; APPLICATION NUMBER: 07/764,465
; FILING DATE: September 23, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/099001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-08-526-136-13

Query Match 8.0%; Score 160; DB 3; Length 466;
Best Local Similarity 21.6%; Pred. No. 6.1e-05;
Matches 84; Conservative 37; Mismatches 90; Indels 178; Gaps 17;

QY 139 SYPPYQATG-PPNTSYMPG-----MPGGISPYPSGYP----- 169
DB 2 SYPGYPTGYPPFGYPAGQESFFPSGQYFYPSPGPPMGGAYPQVPSSGYPGAGYP 61
QY 170 -----FNPSPGPGPYPPGP-YPATSSQ-----YPSQPPVTTVG----- 204
DB 62 APGGYPAPGYPGAPQPGGAPSPYPPGQGFVPPGAGFGSGYP-QPPSQSYGGGPAQV 120
QY 205 -----PSRDGTISE-----DTIRASLI----- 221
DB 121 PLPGGFGGQMPSPGQPYPSQPATVTVQGTIRPAINFDAIRDAEILRKAMKFG 180
QY 222 ---SAVSDKLRWRMKEMDRQAELNAL--KRTEEDLKK----- 255

DB 181 TDSQAIVDVVANRSNDORQIKAAFKTSYKDKLIKLSLSENMELILALFMPPTYD 240
QY 256 -----GHQK---LEEMVTRLDQEAIV-----DKNIELLKKKDEELSSALEKM 295
DB 241 AWSLRKAMQAGTQERVLIILCTRTNQETREIVRCYQSEFGRDLEKDIRSDTSGHFERL 300
QY 296 -----ENQSENNDIDEVIIPITAPLYKQILNLYAENAEATEDTIFYLGEALRRGVID 345
DB 301 LVSMCQGNRDENQSIHQM-----AQEDA--QRLYQAGE-----GRLG 336
QY 346 LDVFLKHVRLLSRKQFOLRALMOKARKTA 374
DB 337 TDESCFNMLLATRSFPOLRATMEAYSMA 365

RESULT 18
US-09-461-325-453
; Sequence 453, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-325-453

Query Match 7.4%; Score 148.5; DB 4; Length 148;
Best Local Similarity 29.7%; Pred. No. 0.00011;
Matches 41; Conservative 18; Mismatches 46; Indels 33; Gaps 6;

QY 126 GDEPPVFSRDISASYPYQATGPPNTSYMPGMPGGISPYPSGYPPNP---SGYPGC--- 178
DB 22 GGSNPAHPPIPNPFPFGPCFP-----PGAPGNPAFPFGPPHPPVPGYPGCQPLG 75
QY 179 PYPPGYPATTSQYPSQPPVTTVGPSPRDTISDITRASLISAVSDKLRWRMKESMDR 238
DB 76 PYPPYPPPPA-----PGIEPVNPLAPGMVGP-----AVIVDK---KMQKXKKK 115
QY 239 AQAEALNALKRTEEDLKKG 256
DB 116 AHKKMHKHQKHKKYKHKG 133

RESULT 19
US-10-012-542-453
; Sequence 453, Application US/10012542
; Patent No. 6627741
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1

```
; CURRENT APPLICATION NUMBER: US/10-012-542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-542-453

Query Match          7.4%; Score 148.5; DB 4; Length 148;
Best Local Similarity 29.7%; Pred. No. 0.00011;
Matches 41; Conservative 18; Mismatches 46; Indels 33; Gaps 6;

QY 126 GDEPPVFRSPISASYPYQATGPPNTSYMPGMPGGISYPSPGPNP---SGYPGC--- 178
DB 22 GGSNPAHPNPINPPFPFGCP---PGAPHGNPAFPFGPHVPQPGYPCQPLG 75

QY 179 PYPGPGYPATSSQYSPQPPVTTVGPRSDGTISEDTIRASLISAVSDKLRWRMKEMDR 238
DB 76 PYPGPYPPPA-----PGIPPNPLAPGMVGP-----AVVDK---KMQKMKK 115

QY 239 AQAEINALKRTEEDLKKG 256
DB 116 AHKMKHKKQHKYKHKG 133

RESULT 20
US-08-574-959A-9
; Sequence 9, Application US/08574959A
; Patent No. 5962224
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi
; APPLICANT: and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/574,959A
; FILING DATE: 19-DEC-95
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
```

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-574-959A-9

Query Match          7.0%; Score 141; DB 2; Length 905;
Best Local Similarity 23.4%; Pred. No. 0.0054;
Matches 68; Conservative 45; Mismatches 140; Indels 38; Gaps 10;

QY 43 MNLFGTIP-VYRGNTYNIPICLMLLDITYPNPPICFVKPTSSMTIKTKHVA----- 95
DB 457 MESAGPVSEPTWTSTANL---LGLLSRPSVCPPLFLFPGPNH---RAGSNEDFLARSG 510
QY 96 -----NGKIVLP-YLHEWKHPQSLGLGIQ-----VMIVVFGDEPPVFSRPI 136
DB 511 TDPPTIPDETFGKVRPAFVHYDKERASDVETLSDDSDSDSVVIVPEGLPLPPPPPS 570
QY 137 SASYPYQATGPPNTSYMPGMPGGISYPSPGPNPSPGYPGCPYPPGGYPATTSSQYPS 196
DB 571 GATPPPIAPTGPPTAS--PPVPA--KEEPEELPAAPGPLPP-PPPPPPVGEVXLPPPPQ 625
QY 197 QPVTTVGPRSDGTISEDTIRASLISAVSDKLRWRMKEMDRAQAEINALKRTEEDLKKG 256
DB 626 LVPEGTGPGGGFPALAEEDLTVININSDDEEEEEESEEEEEESEEEEEESEEEEDF 685
QY 257 HOKLEEMVTRLDQEAEDVKIEILLKKDEELSSALEKMNQSENNDIDEV 307
DB 686 EEEDEDEEYEEEEEEEEEEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFE 735

RESULT 21
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi
; APPLICANT: and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,014
; FILING DATE: 19-Jul-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/574,959
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```



```
QY 96 -----NGKIYLP-YLHEWKHPQSDLLGLIQ-----VMIVFGDEPPVFSRPI 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 741 TPPPTIPPOETFGGRVPRPAFVHYDKKEASDVEISLESDSDSVVIVPEGLPLPLPPPPS 800
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 SASYPYQATGPPNTSYWPGGSGSPSGYPPNPSPGPGPYGPGPYATTSSQVPS 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 801 GATPPPIAPTGFPTAS--PPVDA--KEEFEEFLPAAPGLPP--PPPPPPPVPGVXLPPQ 855
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 QPPVTVGPRDGTISEDTIRASLISAVSDKLRLWMEKMDRAQAEALNALKRTERDLKKG 256
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 856 LVPEGTGGGPPALEEOLTVININSDEEEEGEEEEEDEEEEEEDEEEDEEEDF 915
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 HOKUEMTRLDQVAVDKNIELKKDELUSSALEKMNQSENNDIDEV 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 916 EEEEDDEEYFEEEEEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFE 965
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 24

```
US-09-535-008-63
; Sequence 63, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535, 008
; EARLIER FILING DATE: 2000-03-23
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 543
; TYPE: PRI
; ORGANISM: Homo sapiens
US-09-535-008-63
```

```
Query Match      6.5%; Score 130; DB 4; Length 543;
Best Local Similarity 21.7%; Pred.No.0.021; 71; Indels 122; Gaps 11;
Matches 61; Conservative 27; Mismatches 71; Indels 122; Gaps 11;

QY 126 GDEPVPFSRPIASYPYQATGPPNTSYMPGMP-----GGISPYPSG----- 167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 GPAPENYSRPHGMGGPNMPPGP--SGVPPGMPGQPPGPGPKPWPPEGPMANAAAPTSTPQ 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 --YPPNPSGYGCPYPPGPGP-----YPATTSSQYPSQPP-----VTVGFSR 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 KLIPPQPTGRES-PAPPAVPPFAASVMPPTQSPGQPAQAPMVPFLHOKQSRITPIQKPR 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 -----DGTISED-----TI----- 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GLDPVEILQEREYRLOARIAHRIQLENLPSGLAGDLRTKATIELKALRLNLFQRLRQE 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 217 -----RASLISA-VSDKLRMRKEEMDRA-----QAEALNAL 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 VVCMRRDTALETALNAKAYKRSKQSLREARITKLEKQKQKIEQERKRKRQKHQEYLSI 479
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 KRTEEDLKKGHKLSEEMVTRLDQVAVDKNIELLLKKDEE 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 LQHAQDFKEYHRSVTGKIQLTKAVATYHANTEREQKKENE 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 25

```
US-09-535-008-61
; Sequence 61, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
```

```
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 577
; TYPE: PRI
; ORGANISM: Homo sapiens
US-09-535-008-61

Query Match      6.5%; Score 130; DB 4; Length 577;
Best Local Similarity 21.7%; Pred.No.0.023; 71; Indels 122; Gaps 11;
Matches 61; Conservative 27; Mismatches 27; Indels 122; Gaps 11;

QY 126 GDEPVPFSRPIASYPYQATGPPNTSYMPGMP-----GGISPYPSG----- 167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 GPAPENYSRPHGMGGPNMPPGP--SGVPPGMPGQPPGPGPKPWPPEGPMANAAAPTSTPQ 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 --YPPNPSGYGCPYPPGPGP-----YPATTSSQYPSQPP-----VTVGFSR 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 KLIPPQPTGRES-PAPPAVPPFAASVMPPTQSPGQPAQAPMVPFLHOKQSRITPIQKPR 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 -----DGTISED-----TI----- 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GLDPVEILQEREYRLOARIAHRIQLENLPSGLAGDLRTKATIELKALRLNLFQRLRQE 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 217 -----RASLISA-VSDKLRMRKEEMDRA-----QAEALNAL 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 VVCMRRDTALETALNAKAYKRSKQSLREARITKLEKQKQKIEQERKRKRQKHQEYLSI 479
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 KRTEEDLKKGHKLSEEMVTRLDQVAVDKNIELLLKKDEE 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 LQHAQDFKEYHRSVTGKIQLTKAVATYHANTEREQKKENE 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: July 12, 2004, 08:33:27
Job time : 18.4783 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 10.0131 Seconds
(without alignments)
1981.269 Million cell updates/sec

Title: US-09-804-690-2

Perfect score: 1996
Sequence: 1 MMSKYRDLTVRQTVNVIA.....FQLRALMQKARKTAGLSLDY 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1996	100.0	391	1 T101_MOUSE	Q61187 mus musculus
2	1900.5	95.2	390	1 T101_HUMAN	Q99816 homo sapien
3	293.5	14.7	385	1 ST22 YEAST	P25604 saccharomyc
4	150	7.5	466	1 ANX7_HUMAN	P20073 homo sapien
5	147	7.4	503	1 ANXB_MOUSE	P97384 mus musculus
6	138	6.9	4903	1 MLI3_MOUSE	Q8brh4 mus musculus
7	137.5	6.9	503	1 ANXB_RABIT	P33477 oryctolagus
8	135.5	6.8	463	1 ANX7_MOUSE	Q07076 mus musculus
9	133	6.7	505	1 WASL_BOVIN	Q95107 bos taurus
10	131	6.6	1280	1 DYNA RAT	P28023 rattus norv
11	128	6.4	505	1 ANXB_HUMAN	P50995 homo sapien
12	127.5	6.4	788	1 PCAP_HUMAN	Q96rn5 homo sapien
13	127.5	6.4	1281	1 DYNA_MOUSE	Q08788 mus musculus
14	127	6.4	1224	1 DYNA_CHICK	P35458 gallus gall
15	126.5	6.3	505	1 WASL_HUMAN	Q00401 homo sapien
16	123.5	6.2	375	1 VASP_MOUSE	P70460 mus musculus
17	122.5	6.1	501	1 WASL RAT	Q08816 rattus norv
18	122.5	6.1	503	1 ANXB_BOVIN	P27214 bos taurus
19	122	6.1	978	1 RA50_AQUAE	Q67124 aquifex aeo
20	121.5	6.1	1068	1 DAM2_MOUSE	Q80u19 mus musculus
21	120.5	6.0	338	1 FOSE_HUMAN	P53539 homo sapien
22	119.5	6.0	753	1 ZIN_HUMAN	Q9nr13 homo sapien
23	119.5	6.0	815	1 MK07_HUMAN	Q13164 homo sapien
24	119.5	6.0	1507	1 SET2 CAEEL	Q18221 caenorhabdi
25	119.5	6.0	1733	1 RPB1 YEAST	P04050 saccharomyc
26	119	6.0	379	1 VASP_HUMAN	P50552 homo sapien
27	119	6.0	535	1 Y475_STRMU	Q8dvk7 streptococ
28	119	6.0	792	1 PCAP_MOUSE	Q924h2 mus musculus
29	119	6.0	1068	1 DAM2_HUMAN	Q86t65 homo sapien
30	118.5	5.9	383	1 VASP_CANFA	P50551 canis famil
31	118.5	5.9	559	1 WAS1_HUMAN	Q92558 homo sapien
32	118.5	5.9	760	1 ZIN_MOUSE	P58404 mus musculus
33	118.5	5.9	1847	1 SN24_HUMAN	P51532 homo sapien

ALIGNMENTS

RESULT 1

T101 MOUSE
ID T101 MOUSE STANDARD; PRT; 391 AA.
AC Q61187;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tumor susceptibility gene 101 protein.
GN TSG101.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Fibroblast;
RC MEDLINE=96201522; PubMed=8616888;
RA Li L., Cohen S.N.;
RT "Tsg101: a novel tumor susceptibility gene isolated by controlled
RT homologous functional knockout of allelic loci in mammalian cells.";
RL Cell 85:319-329(1996).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=129/SvJ; TISSUE=Mammary gland;
RC MEDLINE=99054675; PubMed=9840940;
RA Wagner K.-U., Dierissee P., Rucker E.B. III, Robinson G.W.,
RN Hennighausen L.;
RT "Genomic architecture and transcriptional activation of the mouse and
RT human tumor susceptibility gene TSG101: common types of shorter
RT transcripts are true alternative splice variants.";
RL Oncogene 17:2761-2770(1998).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Collins B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RN INTERACTION WITH DMAP1.
RP MEDLINE=20347703; PubMed=10888872;
RA Rountree M.R., Bachman K.E., Baylin S.B.;
RT "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at
RT replication foci";
RL Nat. Genet. 25:269-277(2000).
CC -!- FUNCTION: May be involved in cell growth and differentiation and
CC act as a negative growth regulator.
CC -!- SUBUNIT: Interacts with ubiquitin, statmin and DMAP1.
CC -!- SUBCELLULAR LOCATION: Mainly cytoplasmic. Depending on the stage
CC of the cell cycle, detected in the nucleus.
CC -!- TISSUE SPECIFICITY: Ubiquitous. Higher expression in brain and
CC mammary gland. Lower expression in liver and tumoral tissues.
CC -!- DEVELOPMENTAL STAGE: Expressed at all stages of mammary gland
CC development, but at lower rate at early and mid pregnancy.
CC Expressed in 1-cell and 2-cell stage embryos.
CC -!- DOMAIN: The UEV domain is required for the interaction of the
CC complex with ubiquitin (By similarity).

-!- DOMAIN: The coiled coil domain may interact with statmin.
-!- SIMILARITY: Contains 1 UEV (ubiquitin E2 variant) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcements>
or send an email to license@isb-sib.ch).

EMBL; U52945; AAC53586.1; -;
EMBL; AF060868; AAC83576.1; -;
EMBL; BC005424; AA05424.1; -;
MGD; MGI:106581; Tsg101.
GO; GO:0005515; F:protein binding; IPI.
GO; GO:0007050; P:cell cycle arrest; IDA.
GO; GO:0030216; P:keratinocyte differentiation; IDA.
GO; GO:0008285; P:negative regulation of cell proliferation; IDA.
GO; GO:0001558; P:regulation of cell growth; IDA.
InterPro; IPR008883; Tsg101.
InterPro; IPR000608; UBQ_conjugat.
Pfam; PF05743; Tsg101; 1.
SMART; SM00212; UBCc; 1.
PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; FALSE NEG.
PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; FALSE NEG.
KW Transport; Protein transport; Ubl conjugation pathway; Coiled coil;
KW Growth regulation.
FT DOMAIN 1 133 UEV.
FT DOMAIN 237 317 COILED COIL (POTENTIAL).
SQ SEQUENCE 391 AA; 44123 MW; 79558EC53093492 CRC64;
Query Match 100.0%; Score 1996; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 4e-104;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMSKYYRDLTVRQTVNVIAMYKDLKPLVDSYVFNDSGSSRELNLITGTPVYRGNLYNI 60
Db 11 MMSKYYRDLTVRQTVNVIAMYKDLKPLVDSYVFNDSGSSRELNLITGTPVYRGNLYNI 70
QY 61 PICLWLDDTYPNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLDHDKHPRSELLELIQI 120
Db 71 PICLWLDDTYPNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLDHDKHPRSELLELIQI 130
QY 121 MIVIFGEPPVFSRPVSASYPPTATGTPNTSYMPGMPGSIAYSPGPNPSGYPGCP 180
Db 131 MIVIFGEPPVFSRPVSASYPPTATGTPNTSYMPGMPGSIAYSPGPNPSGYPGCP 190
QY 181 YPPAGYPATTSSQYPSQPPVTVGFSRDGTISDITRASLISAVSDKLRWRKKEWDGA 240
Db 191 YPPAGYPATTSSQYPSQPPVTVGFSRDGTISDITRASLISAVSDKLRWRKKEWDGA 250
QY 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNTIELLKKKDEELSSALEKMNQS 300
Db 251 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNTIELLKKKDEELSSALEKMNQS 310
QY 301 ENNDIDEVIPTAPLYKQILNLYAEENAIETDTFYLGALRRGVLDLDFLKHVRLSRK 360
Db 311 ENNDIDEVIPTAPLYKQILNLYAEENAIETDTFYLGALRRGVLDLDFLKHVRLSRK 370
RESULT 2
T101_HUMAN
ID T101_HUMAN STANDARD; PRT; 390 AA.
AC Q99816;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tumor susceptibility gene 101 protein.

```
GN TSG101.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=97148696; PubMed=9019400;
RA Li L., Li X., Francke U., Cohen S.N.;
RT "The TSG101 tumor susceptibility gene is located in chromosome 11 band
RL p15 and is mutated in human breast cancer.";
RN Cell 88:143-154(1997).
RN [2]
RN ERATUM.
RP MEDLINE=99080983; PubMed=9867424;
RA Li L., Francke U., Cohen S.N.;
RN Cell 93:660-660(1998).
RN [3]
RN ALTERNATIVE SPLICING (ISOFORM 2).
RP MEDLINE=98031856; PubMed=9366528;
RA Gayther S.A., Barski P., Batley S.J., Li L., de Foy K.A., Cohen S.N.,
RA Ponder B.A., Caldas C.;
RT "Aberrant splicing of the TSG101 and FHIT genes occurs frequently in
RL multiple malignancies and in normal tissues and mimics alterations
RL previously described in tumours.";
RN Oncogene 15:2119-2126(1997).
RN [4]
RN ALTERNATIVE SPLICING.
RP MEDLINE=97384943; PubMed=9242438;
RA Lee M.P., Feinberg A.P.;
RT "Aberrant splicing but not mutations of TSG101 in human breast
RL cancer.";
RN Cancer Res. 57:3131-3134(1997).
RN [5]
RN ALTERNATIVE SPLICING.
RP MEDLINE=99054675; PubMed=9840940;
RA Wagner K.-U., Dierisseau P., Rucker E.B. III, Robinson G.W.,
RA Hennighausen L.;
RT "Genomic architecture and transcriptional activation of the mouse and
RL human tumor susceptibility gene TSG101: common types of shorter
RL transcripts are true alternative splice variants.";
RN Oncogene 17:2761-2770(1998).
RN [6]
RN INTERACTION WITH DMAP1.
RP MEDLINE=20347709; PubMed=10888872;
RA Rountree M.R., Bachman K.E., Baylin S.B.;
RT "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at
RL replication foci.";
RN Nat. Genet. 25:269-277(2000).
RN [7]
RN STRUCTURE BY NMR OF 1-145.
RP MEDLINE=22001147; PubMed=12006492;
RA Pornillos O., Alam S.L., Rich R.L., Myszka D.G., Davis D.R.,
RA Pornillos W.I.;
RT "Structure and functional interactions of the Tsg101 UEV domain.";
RN EMBO J. 21:2397-2406(2002).
RN [8]
RN STRUCTURE BY NMR OF 1-145.
RP MEDLINE=22289677; PubMed=12379843;
RA Pornillos O., Alam S.L., Davis D.R., Sundquist W.I.;
RT "Structure of the Tsg101 UEV domain in complex with the PTAP motif of
RL the HIV-1 p6 protein.";
RN Nat. Struct. Biol. 9:812-817(2002).
RN [9]
RN FUNCTION: May be involved in cell growth and differentiation and
CC act as a negative growth regulator.
CC -! SUBUNIT: Interacts with ubiquitin, stathmin and DMAP1.
CC -! SUBCELLULAR LOCATION: Mainly cytoplasmic. Depending on the stage
CC of the cell cycle, detected in the nucleus.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist. Several shorter
CC isoforms are detected in primary breast cancers and other
```

```

Db          370 QFQRLALMQARKTAGLSDLV 390
|||||
RESULT 3
ST22 YEAST
ID ST22 YEAST STANDARD; PRT; 385 AA.
AC P25604; P87279; Q862T3; Q8NIM6;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Suppressor protein STP22 of temperature-sensitive alpha-factor
DE receptor and arginine permease (Vacuolar protein sorting-associated
DE protein VPS23).
DN STP22 OR VPS23 OR YCL008C OR YCL8C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RP MEDLINE=99223587; PubMed=10207082;
RA Li Y., Kane T., Tipper C., Spatrick P., Jenness D.D.;
RT "Yeast mutants affecting possible quality control of plasma membrane
RT proteins."
RL Mol. Cell. Biol. 19:3588-3599(1999).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=S288c;
RX MEDLINE=92244356; PubMed=1574125;
RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA Carignani G., Chanet R., Contreras R., Crouzet M., Daiguan-Fornier B.,
RA De Haan B., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA Dujon B., Duesterhoef A., Erdmann D., Esteban M., Fabre F.,
RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA Francinques-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glandsdorff N.,
RA Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haaseemann M.,
RA Hatat D., Hegeman J.H., Herbet C.J., Hilger F., Hohmann S.,
RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackson P.,
RA Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA Kleinhaus U., Kreis P., Lafranchi G., Lewis C., van der Linden C.G.,
RA Lucchini G., Lutzenkirchen K., Maat C., Mannheim G., Manzano M.E.,
RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA Perea J., Philippson P., Pierard A., Planta R.J., Plevani P.,
RA Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA Raynal A., Remacha M., Richerich P., Roberts A.B., Rodriguez F.,
RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA Spiegelberg R., Staveva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA Thireos G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,
RA van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warming J.R.,
RA von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RA Zimmermann F.K., Sgourou J.G.;
RT "The complete DNA sequence of yeast chromosome III."
RL Nature 357:38-46(1992).
RN [3]
REVSIONS.
RP Gromadka R.;
RA Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
REVSIONS TO N-TERMINUS.
RP Valles G., Volckaerts G.;
RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
SEQUENCE OF 204-329 FROM N.A.
RP STRAIN=S288c;
RX MEDLINE=22728591; PubMed=12844361;

```

```

RA Brachat S., Dietrich F.S., Voegeli S., Zhang Z., Stuart L., Lerch A.,
RA Gates K., Gaffney T., Philippen P.;
RT "Reinvestigation of the Saccharomycetes cerevisiae genome annotation by
RT comparison to the genome of a related fungus: Ashbya gossypii."
RL Genome Biol. 4:RESEARCH45.1-RESEARCH45.13(2003).
RN [6]
FUNCTION, AND SUBCELLULAR LOCATION.
RP MEDLINE=21129038; PubMed=11208108;
RA Babst M., Odorizzi G., Estepa E.J., Emr S.D.;
RT "Mammalian tumor susceptibility gene 101 (TSG101) and the yeast
RT homologue, Vps23p, both function in late endosomal trafficking."
RL Traffic 1:248-258(2000).
RN [7]
FUNCTION, SUBUNITS, SUBCELLULAR LOCATION, AND MUTAGENESIS OF MET-85.
RP MEDLINE=21402413; PubMed=11511343;
RA Katzmann D.J., Babst M., Emr S.D.;
RT "Ubiquitin-dependent sorting into the multivesicular body pathway
RT requires the function of a conserved endosomal protein sorting
RT complex, ESCRT-I."
RL Cell 106:145-155(2001).
RN [8]
FUNCTION: The ESCRT-I complex recognizes ubiquitinated
CC multivesicular body (MVB) cargo. It is required for sorting into
CC MVB vesicles. Required for vacuolar targeting of temperature-
CC sensitive plasma membrane proteins STP2 and CAN1.
CC -I- SUBUNIT: Component of ESCRT-I, endosomal sorting complex required
CC for transport I, which consists of STP22, VPS28 and SRN2.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic and endosomal.
CC -I- DOMAIN: The UEV domain is required for the interaction of the
CC complex with ubiquitin.
CC -I- SIMILARITY: Contains 1 UEV (ubiquitin E2 variant) domain.
CC -I- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 294.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004731; AAB62820.1; -
CC EMBL; X59720; CAC42964.1; ALT_FRAME.
CC EMBL; AY260880; AAP21748.1; -
CC PIR; S74288; S74288.
CC Germonline; 138851; -
CC SGD; S0000514; STP22.
CC GO; GO:0005768; C:endosome; IDA.
CC GO; GO:0005515; F:protein binding; IDA.
CC GO; GO:0006612; P:protein-membrane targeting; IMP.
CC GO; GO:0006623; P:protein-vacuolar targeting; IMP.
CC InterPro; IPR008883; Tsg101.
CC InterPro; IPR000608; UBQ_conjugat.
CC Pfam; PF05743; Tsg101; 1.
CC SMART; SM00212; UBCC; 1.
CC PROSITE; PS00183; UBIQUITIN_CONUGAT_1; FALSE_NEG.
CC PROSITE; PS0127; UBIQUITIN_CONUGAT_2; 1.
CC Transport; Protein transport; Ubl conjugation pathway; Coiled coil.
FT DOMAIN 87 164
FT DOMAIN 155 201
FT DOMAIN 272 300
FT MUTAGEN 85 85 M->T: NO INTERACTION OF THE ESCRT-I
FT COMPLEX WITH UBIQUITIN.
FT SEQUENCE 385 AA; 43330 MW; FEDE3BE79F063BCE CRC64;
Query Match 14.7%; Score 293.5; DB 1; Length 385;
Best Local Similarity 26.0%; Pred. No. 1.3e-09;
Matches 102; Conservative 74; Mismatches 164; Indels 53; Gaps 16;
QY 7 YRD--LTVRQTVNVIAMKYKDLKPVLDVYVFNDSGSSRELNVLTCTIPVRVGNL--YNIPIC 63
DB 25 YNDGRTHFDHSDALLDNFHSLRPRTRVTHSDGTPQLLSIYGTISTGDSGSSPHSIPVI 84

```

[illegible]

RESULT 4
ANX7_HUMAN
ID ANX7 HUMAN STANDARD; PRT; 466 AA.
AC P20073;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Annexin A7 (Annexin VII) (Synexin).
GN ANXA7 OR ANX7
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=89264510; PubMed=2542947;
RX Burns A.L., Magendzo K., Shirvan A., Srivastava M., Rojas E.,
RA Aljani M.R., Pollard H.B.;
RA "Calcium channel activity of purified human synexin and structure of
RT the human synexin gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3798-3802(1989).
[2]
RN SEQUENCE FROM N.A.
RP

RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavet T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallada D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buttefield Y.S.N., Kyrzyanski M.I., Skalski J., Smalus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- FUNCTION: Calcium/phospholipid-binding protein which promotes
 CC membrane fusion and is involved in exocytosis.
 CC -!- DOMAIN: A pair of annexin repeats may form one binding site for
 CC

RESULT 5
ANXB_MOUSE
ID ANXI
AC P97:

CC	calcium and phospholipid.
CC	-!- SIMILARITY: Belongs to the annexin family.
CC	-!- SIMILARITY: Contains 4 annexin repeats.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; J04543; AAA36616.1; -.
DR	EMBL; BC002632; AAH02632.1; -.
DR	PIR; A54467; LUHU7.
DR	HSSP; P26256; LDMS.
DR	Genew; HGNC:545; ANXA7.
DR	MIM; 186360; -.
DR	InterPro; IPR001464; Annexin.
DR	pfam; PF00191; annexin; 4.
DR	PRINTS; PR00196; ANNEXIN.
DR	ProDom; PD000143; Annexin; 4.
DR	SMART; SM00335; ANX; 4.
DR	PROSITE; PS00223; ANNEXIN; 4.
DR	Annexin; Calcium/phospholipid-binding; Repeat.
DR	DOMAIN 1 143 REPEAT-RICH REGION.
FT	REPEAT 172 232 ANNEXIN 1.
FT	REPEAT 244 304 ANNEXIN 2.
FT	REPEAT 327 387 ANNEXIN 3.
FT	REPEAT 403 463 ANNEXIN 4.
FT	DOMAIN 5 20 3 X 5 AA TANDEM REPEATS OF G-Y-P-P-X.
FT	REPEAT 5 9 1.
FT	REPEAT 10 14 2.
FT	REPEAT 16 20 3.
FT	SEQUENCE 466 AA; 50315 MW; 09A6760729D45FCD CRC64;
SO	

```

RESULT 5
ANXB MOUSE
ID _ANXB MOUSE
AC P97384;
STANDARD;
PRT; 503 AA.

```

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
 GN ANXA11 OR ANXA11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97092887; PubMed=8938449;
 RA Fernandez M.P., Jenkins N.A., Gilbert D.J., Copeland N.G.,
 RA Morgan R.O.;
 RT "Sequence and chromosomal localization of mouse annexin XI.";
 RL Genomics 37:366-374 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=129/SVJ;
 RX MEDLINE=20469408; PubMed=11013079;
 RA Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
 RA Fernandez M.P.;
 RT "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
 annexins and source of orthologous cDNA isoforms.";
 RL Genomics 69:95-103 (2000).
 CC -1- FUNCTION: Binds specifically to calyculin in a calcium-dependent
 manner.
 CC -1- DOMAIN: A pair of annexin repeats may form one binding site for
 calcium and phospholipid.
 CC -1- SIMILARITY: Belongs to the annexin family.
 CC -1- SIMILARITY: Contains 4 annexin repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U65986; AAB42012.1; -.
 DR EMBL; AJ289760; CAB94770.1; -.
 DR EMBL; AJ289761; CAB94770.1; JOINED.
 DR EMBL; AJ289762; CAB94770.1; JOINED.
 DR EMBL; AJ289763; CAB94770.1; JOINED.
 DR EMBL; AJ289764; CAB94770.1; JOINED.
 DR EMBL; AJ289765; CAB94770.1; JOINED.
 DR EMBL; AJ289766; CAB94770.1; JOINED.
 DR EMBL; AJ289767; CAB94770.1; JOINED.
 DR EMBL; AJ289768; CAB94770.1; JOINED.
 DR EMBL; AJ289769; CAB94770.1; JOINED.
 DR HSSP; P13214; IANN
 DR SWISS-2DPAGE; P97384; MOUSE.
 DR MGD; MGI:108481; Anx11.
 DR GO; GO:0005635; C:nuclear membrane; ISS.
 DR GO; GO:0005654; C:nucleoplasm; ISS.
 DR GO; GO:0005515; F:protein binding; ISS.
 DR InterPro; IPR001464; Annexin.
 DR Pfam; PF00191; annexin; 4.
 DR PRINTS; PR00196; ANNEXIN.
 DR ProDom; PD000143; Annexin.
 DR SMART; SM00335; ANX; 4.
 DR PROSITE; PS00223; ANNEXIN; 4.
 DR Annexin; Calcium/phospholipid-binding; Repeat.
 FT REPEAT 207 267 ANNEXIN 1.
 FT REPEAT 279 339 ANNEXIN 2.
 FT REPEAT 363 423 ANNEXIN 3.
 FT REPEAT 438 498 ANNEXIN 4.
 SQ SEQUENCE 503 AA; 54111 MW; 424B1345E0F4EC8A CRC64;
 Query Match 7.4%; Score 147; DB 1; Length 503;
 Best Local Similarity 26.3%; Pred. No. 0.24; Indels 92; Gaps 16;
 Matches 77; Conservative 29; Mismatches 95;

QY 125 FGEEPPVFSRPTVSASYPPTATGPPNTSNWPMGSPGISAYPSGYPNPSGYPGCPVPPA 184
 DB 89 FGQPPPA-QQPV-----PPYGMVPPPGGPPGMPG-----YPA-YPGAP--VPQPMPT 135
 QY 185 GPYPATTSOYPSQPPVTVG-----PSRDGTISEDITRASL-----I 222
 DB 136 GQQP-----PGAYPGQPPMTYFGQSPMPPPGQOPVPSYFGSGSSTITPAVPPAQFGNRTI 192
 QY 223 SAVS-----DKLRMRKEEMDGAQALNALKRTEEDLKKGHQKLEEMVTRLDQVAE 274
 DB 193 TAASGFDPRLDAEVLKAMKGFQDEQAIDCLGSRN--KQOQILLSKFYAKGK----- 246
 QY 275 VDKNIELKKKDELSALEKMNQSNNDIDEVIPTAPLYKQILMLYAEENAIETIF 334
 DB 247 -----DLIKDLKSELGSGNFE-----KTILALM--KTPVLEVDY 277
 QY 335 YLGEALRGVID---LDVFL---KVRLLSRKQFOLRALMOKARKTAGLSD 379
 DB 278 EIKKAIKAGTDEACLEIFASRSNEHIRELSR---AYKTEFQKTLLEAIRSD 327

RESULT 6

MLL3 MOUSE
 ID_MLL3_MOUSE STANDARD; PRT; 4903 AA.
 AC QBRHR4; Q8BK12; Q8C6M3; Q923H5; Q923H6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-
 DE lysine N-methyltransferase, H3 lysine-4 specific MLL3) (EC 2.1.1.43).
 GN MLL3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP CONCEPTUAL TRANSLATION OF 814-956 AND 1377-4213.
 RA Blatter M.-C.;
 RL Unpublished observations (SEP-2003).
 RN [2]
 RP SEQUENCE OF 957-1376 AND 4214-4894 FROM N.A.
 RC TISSUE=Myeloma;
 RX MEDLINE=21574953; PubMed=11718452;
 RA Tan Y.C., Chow V.T.;
 RT "Novel human HALL (MLL3) gene encodes a protein homologous to ALR and
 RT to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated
 RT with leukemia and developmental defects.";
 RL Cancer Detect. Prev. 25:454-469 (2001).
 RN [3]
 RP SEQUENCE OF 1-813 AND 4803-4903 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt B., Brusci V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konegaya A., Kurochkin I.V., Lee Y., Lenhard R., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempole C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,

RA Hirozane-Kishikawa T., Komno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a
CC coactivator complex of nuclear receptors, involved in
CC transcriptional coactivation. MLL3 may be a catalytic subunit of
CC this complex, which weakly methylates Lys-4 of histone H3. This is
CC a specific tag for epigenetic transcriptional activation (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins MLL2
CC and MLL3, and ASH2/ASCL2. Interacts with histone H3 (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B
CC and H4, and may have a H3 lysine specific methylation activity (By
CC similarity).
CC -!- SIMILARITY: Belongs to the transcription factor trithorax family.
CC -!- SIMILARITY: Contains 1 DHHC-type zinc finger.
CC -!- SIMILARITY: Contains 6 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- CAUTION: The regions from 814 to 956 and 1377 to 4213 were
CC deduced from the genomic sequence by similarity to the human
CC sequence.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AY036886; AAK70213.1; -;
DR EMBL; AY036887; AAK70214.1; -;
DR EMBL; AK044828; BAC32109.1; -;
DR EMBL; AK054270; BAC35712.1; -;
DR EMBL; AK077567; BAC36867.1; -;
DR MGD; MGI:2444959; Mll3.
DR InterPro; IPR000637; At hook.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF02178; At_hook; 1.
DR Pfam; PF00628; PHD; 2.
DR SMART; SM00249; PHD; 3.
DR PROSITE; PS00354; HMG1_Y; 1.
DR PROSITE; PS00868; POST_SET; 1.
DR PROSITE; PS0280; SET_1.
DR PROSITE; PS0216; ZF_DHC; 1.
DR PROSITE; PS01359; ZF_PHD_1; 5.
DR PROSITE; PS0016; ZF_PHD_2; 6.
DR PROSITE; PS0089; ZF_RING_2; 1.
KW Transferase; Methyltransferase; Chromatin regulator; Activator;
KW DNA-binding; Nuclear protein; Transcription regulation; Coiled coil;
KW Zinc-finger; Repeat.
FT ZN_FING 340 390 PHD-TYPE 1.
FT RING-TYPE 1.
FT ZN_FING 343 388 RING-TYPE 2.
FT ZN_FING 387 437 DHHC-TYPE 2.
FT ZN_FING 435 488 DHHC-TYPE 3.
FT ZN_FING 463 519 PHD-TYPE 3.
FT ZN_FING 950 1003 PHD-TYPE 4.
FT ZN_FING 1000 1050 PHD-TYPE 5.

FT ZN_FING 1077 1132 PHD-TYPE 6.
FT DNA_BIND 34 46 A.T HOOK (BY SIMILARITY).
FT DOMAIN 4762 4883 SET.
FT DOMAIN 4887 4903 POST-SET.
FT DOMAIN 1330 1352 COILED COIL (POTENTIAL).
FT DOMAIN 1352 1790 COILED COIL (POTENTIAL).
FT DOMAIN 1743 3074 COILED COIL (POTENTIAL).
FT DOMAIN 3166 3193 COILED COIL (POTENTIAL).
FT DOMAIN 3224 3270 COILED COIL (POTENTIAL).
FT DOMAIN 3387 3432 COILED COIL (POTENTIAL).
FT DOMAIN 970 1099 CVS-RICH.
FT DOMAIN 284 487 CVS-RICH.
FT DOMAIN 1519 1568 PRO-RICH.
FT DOMAIN 1708 1787 GLN-RICH.
FT DOMAIN 1831 2622 PRO-RICH.
FT DOMAIN 2682 2780 ASP-RICH.
FT DOMAIN 3022 3504 GLN-RICH.
FT CONFLICT 433 NCR1C1EC -> VSDFLICF (IN REF. 3;
FT CONFLICT 675 BAC32109).
FT CONFLICT 675 S -> C (IN REF. 3; BAC35712).
FT CONFLICT 4713 MISSING (IN REF. 2; AAK70214).
SQ SEQUENCE 4903 AA; 540182 MW; 0B896490B081BA6C CRC64;

Query Match 6.9%; Score 138; DB 1; Length 4903;
Best Local Similarity 22.1%; Pred. No. 11; Indels 106; Gaps 16;
Matches 79; Conservative 42; Mismatches 130;

QY 26 KPVLDYVFNDDGSSRELNVNLGTIPVRYRGNIYINIPICLWLLDTYPYNPPICFVKPTSSM 85
DB 2132 RPLIDSYSQTSARSNDPPYSPQPTPRN-----TIDPYSQQPTPRSPQTDN 2182
QY 86 ---TTKTKGHVDANGKIYLPYLHDKWKRSELLELIQIMVIFGEBPPVFSPTVSASVP 142
DB 2183 FVSSVANQRHTD-----PYTHLGLPRPG-----ISVPYQPAPVPRPTSEGT 2227
QY 143 PYATGCP---PN-----TSYMPGMSGISAYPS--GYPPNPGYRG----- 178
DB 2228 RPSSARPALMPDQDFLQAQNRVFLGFLPLRPDPTCSQTPRPG-PGRITFTTHASS 2286
QY 179 ---CPY--PPAGFYPATTS---SQY-----PSQPPVTVTG----- 205
DB 2287 AVRDIYDQPPVTPRPHSESGFTSQVVDLVRPVGSEGNFTSSNLPVSSQSQPFSSVS 2346
QY 206 -----PSRDGTISDTRASLISAVSKLVRWKEEMDGAELNALKRTEDLKKGHQ 259
DB 2347 QLPGVPVTSGGTDTQNTVNMS--QADTEKLQRQK-----LREIILQQQKKIAS 2395
QY 260 KLEEMVTRLDQEAQVVDKNIKLLKKDELSALEKMNQSENNDIVIIPTAPLY 316
DB 2396 ROE----KGQDTAVVPHVPLPHWQPSINQAFTRPPPPYPGSTRSPVPLGPY 2448

RESULT 7
ANXB RABIT
ID ANXB RABIT STANDARD; PRT; 503 AA.
AC F33477;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
GN ANX11 OR ANX11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=92378579; PubMed=1380798;
RA Tokumitsu H., Mizutani A., Muramatsu M.-A., Yokota T., Arai K.-I.,
RA Hidaka H.;
RT "Molecular cloning of rabbit CAP-50, a calcyclin-associated annexin
RT protein.";


```
Db 323 RLYQAGE-----GRLGTDESCFNMLATRSFPOLKATMEAYSMA 362
RESULT 9
WASL_BOVIN
ID WASL_BOVIN STANDARD; PRT; 505 AA.
AC Q95107;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neural Wiskott-Aldrich syndrome protein (N-WASP).
GN WASL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97050838; PubMed=8895577;
RA Miki H., Miura K., Takenawa T.;
RT "N-WASP, a novel actin-depolymerizing protein, regulates the cortical
cytoskeletal rearrangement in a PIP2-dependent manner downstream of
tyrosine kinases.";
RT cytoskeletal rearrangement in a PIP2-dependent manner downstream of
RL EMBO J. 15:5326-5335(1996).
CC -!- FUNCTION: REGULATES ACTIN POLYMERIZATION BY STIMULATING THE ACTIN-
NUCLEATING ACTIVITY OF THE ACTIN-RELATED PROTEIN 2/3 (ARP2/3)
COMPLEX.
CC -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX; INTERACTS WITH CDC42
CC -!- BINDS TO SH3 DOMAINS OF ASH/GRB2.
CC -!- SIMILARITY: Contains 1 CRIB domain.
CC -!- SIMILARITY: Contains 1 WH1 domain.
CC -!- SIMILARITY: Contains 2 WH2 domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; D67066; BAA11082.1; -.
DR PIR; S72273; S72273.
DR InterPro; IPR000697; EVH1.
DR InterPro; IPR000095; PAKbox/RhoGndg.
DR InterPro; IPR001960; WH1.
DR InterPro; IPR003124; WH2.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF00568; WH1; 1.
DR Pfam; PF02205; WH2; 2.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00461; WH1; 1.
DR SMART; SM00246; WH2; 2.
DR PROSITE; PS0108; CRIB; 1.
KW Actin-binding; Repeat.
FT DOMAIN 34 138 WH1.
FT DOMAIN 203 216 CRIB.
FT DOMAIN 277 392 PRO-RICH.
FT DOMAIN 405 422 WH2 1.
FT DOMAIN 433 450 WH2 2.
FT DOMAIN 486 505 ASP-RICH.
SQ SEQUENCE 505 AA; 54671 MW; 54B83B48F1CDB3B8 CRC64;

Query Match 6.7%; Score 133; DB 1; Length 505;
Best Local Similarity 27.0%; Pred. No. 1.5;
Matches 58; Conservative 31; Mismatches 66; Indels 60; Gaps 12;

QY 126 GEPPTVFESR-PTVSASVPPYATG-----PPNTSYMPGMSGISAYPSGYPNP----- 173
Db 308 GAPPPTPSRAPTAAPPPPPPSRGVGAPPPPPNMYPPPLPALPSSAPSGPPPPPLSV 367
```

```
QY 174 SGYPCPYPPAGPYPATYSSQYPSQPPVTVTPSRDGTISEDTIRASLISAVSDKLRWM 233
Db 368 SGSVAPPPPPPPPPPP-----PGPPP--PPGLPSGDGHQVPTPAGSK-AALLDQIR--- 414
QY 234 KEEMDGAQALNALKRTEDLK-----KGHQKLEEMVTRLQ--EVAEVDKN----- 278
Db 415 -----EGAQ-----LKKVEQNSRPVSCGRDALLDQIRQIQIKSVTDPESTPPAPAPTS 465
QY 279 -----IELKKKBELSALEKMQNSNNIDE 307
Db 466 GIVGALMEVMQKRSKAIHSS-----DEDEDEDDDE 495

RESULT 10
DYNA RAT
ID DYNA RAT STANDARD; PRT; 1280 AA.
AC P28023;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE (p150-glued).
GN DCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=91260877; PubMed=1828535;
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitt N.G.,
Pfister K.K., Vallee R.B.;
RT "Homology of a 150K cytoplasmic dynein-associated polypeptide with
the Drosophila gene Glued.";
RL Nature 351:579-580(1991).
RN [2]
RP REVISIONS.
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitt N.G.,
Pfister K.K., Vallee R.B.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required for the cytoplasmic dynein-driven retrograde
movement of vesicles and organelles along microtubules. Dynein-
dynactin interaction is a key component of the mechanism of axonal
transport of vesicles and organelles.
CC -!- SUBUNIT: Large macromolecular complex of at least 10 components;
p150(glued) binds directly to microtubules and to cytoplasmic
dynein.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the dynactin 150 kDa subunit family.
CC -!- SIMILARITY: Contains 1 CAP-Gly domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X62160; CAA44091.1; -.
DR PIR; S16129; S16129.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_GLY; 1.
DR PROSITE; PS00845; CAP_GLY_1; 1.
DR PROSITE; PS50245; CAP_GLY_2; 1.
KW Motor protein; Microtubule; Dynein; Coiled coil; Cytoskeleton.
FT DOMAIN 48 90 CAP-GLY.
FT DOMAIN 157 184 SER-RICH.
FT DOMAIN 214 513 COILED COIL (POTENTIAL).
FT DOMAIN 942 1048 COILED COIL (POTENTIAL).
FT DOMAIN 1184 1213 COILED COIL (POTENTIAL).
SQ SEQUENCE 1280 AA; 141929 MW; C9348CF129FAFF5C CRC64;
```



```
Db          249 DLIKDKSELSGNFEK 264
:|:|      |||  ||
:|:|      |||  ||

RESULT 12
PCAP HUMAN
ID PCAP HUMAN STANDARD; PRT; 788 AA.
AC Q96RNS; O15413; Q8NF16; Q96CT0; Q96IH7; Q9PIT3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Positive cofactor 2 glutamine/Q-rich-associated protein (PC2
DE glutamine/Q-rich-associated protein) (TPA-inducible gene-1) (TIG-1)
DE (Activator-recruited cofactor 105 kDa component) (ARCI105) (CTG repeat
DE protein 7a).
GN PCOAP OR TIG1 OR ARCI105 OR CTG7A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Megakaryocytes, and Placenta;
RX MEDLINE=20480707; PubMed=11024300;
RA Abraham S., Solomon W.B.;
RT "A novel glutamine-rich putative transcriptional adaptor protein
RT (TIG-1), preferentially expressed in placental and bone-marrow
RT tissues."
RL Gene 255:389-400(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND POLYMORPHISM OF POLY-GLN REGION.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=21309066; PubMed=11414760;
RA Berti L., Mittler G., Przemeck G.K.H., Stelzer G., Guenzler B.,
RA Amati F., Conti E., Dallapiccola B., Hrabe' de Angelis M., Novelli G.,
RA Meisternst M.;
RT "Isolation and characterization of a novel gene from the DiGeorge
RT chromosomal region that encodes for a mediator subunit."
RL Genomics 74:320-332(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Hepatosoma, and Spleen;
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ora T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isoqai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Eye, Kidney, and Uterus;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altschuler R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 185-573 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain cortex;
RX MEDLINE=97369492; PubMed=9225980;
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,
RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT "cDNAs with long CAG trinucleotide repeats from human brain."
RL Hum. Genet. 100:114-122(1997).
RN [6]
RP IDENTIFICATION IN ARC COMPLEX, AND SEQUENCE OF 39-48 AND 525-536.
RX MEDLINE=99249346; PubMed=10235267;
RA Naeae A.M., Beaurang P.A., Zhou S., Abraham S., Solomon W.B.,
RA Tian R.;
RT "Composite co-activator ARC mediates chromatin-directed
RT transcriptional activation."
RL Nature 398:828-832(1999).
CC -!- FUNCTION: May function as a transcriptional coactivator in RNA
CC polymerase II transcription. As a protein complex component it may
CC regulate gene transcription, especially following induction of
CC protein kinase C activity.
CC -!- SUBUNIT: Subunit of the large multiprotein complexes PC2 and
CC ARC/DRIP.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q96RNS-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q96RNS-2; Sequence=VSP_003922;
CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined, including
CC heart, brain, lung, spleen, thymus, pancreas, blood leukocyte and
CC placenta. However, the level of expression varied, with highest
CC expression in the placenta and peripheral blood and lowest in the
CC pancreas and kidney.
CC -!- INDUCTION: By 12-O-tetradecanoylphorbol-13-acetate (TPA).
CC -!- POLYMORPHISM: The poly-Gln region from amino acids 235-262 of
CC PCOAP is polymorphic. There are from 15 to 18 repeats in the
CC Italian population.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 13, 600 and 749.
CC -!- CAUTION: Ref.3 (BAB85034) sequence differs from that shown due to
CC sequencing errors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF056191; AAC12944.1; ALT_FRAME.
CC EMBL; AF328769; AAK58423.1; -.
CC EMBL; AK074268; BAB85034.1; ALT_SEQ.
CC EMBL; AK090465; BAC03446.1; ALT_INIT.
CC EMBL; BC007529; AAH07529.1; -.
CC EMBL; BC013985; AAH13985.1; -.
CC EMBL; BC017110; AAH17110.1; ALT_INIT.
CC EMBL; U080745; AAB91443.1; -.
CC Genew; HGNC:14248; PCOAP.
CC MIM; 607372; -.
CC KW Transcription regulation; Activator; Nuclear protein; Polymorphism;
CC Triplet repeat expansion; Alternative splicing.
FT DOMAIN 547 584 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 161 174 POLY-GLN.
FT DOMAIN 178 193 POLY-GLN.
FT DOMAIN 205 218 POLY-GLN.
FT DOMAIN 226 239 POLY-GLN.
FT DOMAIN 243 262 PRO-RICH.
FT DOMAIN 266 315 PRO-GLN.
FT DOMAIN 360 367 POLY-GLN.
FT DOMAIN 449 456 POLY-PRO.
FT DOMAIN 602 611 Missing (in isoform 2).
FT VARSPLOC 385 424 /FTId=VSP_003922.
```

```

FT VARIANT 261 262 Missing.
FT CONFLICT 12 12 S -> R (IN REF. 1).
FT CONFLICT 116 116 L -> F (IN REF. 1).
FT CONFLICT 154 154 Q -> H (IN REF. 3; BAC03446).
FT CONFLICT 161 161 Q -> R (IN REF. 3; BAB85034).
FT CONFLICT 185 186 QQ -> EL (IN REF. 5).
FT CONFLICT 232 287 MISSING (IN REF. 3; BAB85034).
FT CONFLICT 265 265 Q -> E (IN REF. 1 AND 5).
FT CONFLICT 572 573 IL -> GI (IN REF. 3).
FT CONFLICT 685 685 L -> V (IN REF. 3).
SQ SEQUENCE 788 AA; 86753 MW; BB6AC6C63ED2F97E CRC64;

Query Match 6.4%; Score 127.5; DB 1; Length 788;
Best Local Similarity 23.4%; Pred. No. 5;
Matches 63; Conservative 31; Mismatches 74; Indels 101; Gaps 15;

QY 90 GKHDVANGKIYLYLHDKWKRSELLELIQIMVIFGEEP-----PVFSRPT----- 136
DB 394 GMHRA-----RPPPTTAVSAIPSSIPLGRQPMAQVQSLSPLMLSSPSPGQQ 441

QY 137 --VSASYPPYATGPPNTSYMPGMP-----SGISAYPSGYPPNPSPGCPYPAG 185
DB 442 VQTQSMPP-----PPQSPQPGQSPQSPNSVNSGAPSPSPFLSPS-----PQSPQS 491

QY 186 PYPATTSQY--PSOPPVT--VGPSRDGTISDITRASLISAVSDKLRWRMKEEMDGAQA 242
DB 492 PVYARTQNSVSPSGPLNTFVNES-----SVMSPASS-----QAEE-----QQ 531

QY 243 ELNALKTEBDLKGKHKLEMTVRLDQEAQVAVDKNIELKKKD----- 286
DB 532 YLDKIKQ---LSKYIEPLRRMINKIDK--NEDRKDKLSKMSKLLDILTDPKRCPLKT 584

QY 287 -EELSSALEKMNSENNDIDEVIIPTAP 314
DB 585 LQCEIALEKLN-----DMAVPTPP 605

RESULT 13
DINA MOUSE
ID DINA_MOUSE STANDARD; PRT; 1281 AA.
AC O08788;
DT 15-JUL-1998 (Rel. 36, Last Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE (p150-glued).
GN DCTN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=97223454; PubMed=9070275;
RA Jang W., Weber J.S., Tokito M.K., Holzbaur E.L., Meisler M.H.;
RT "Mouse p150Glued (dynactin 1) cDNA sequence and evaluation as a
RT candidate for the neuromuscular disease mutation mnd2."
RL Biochem. Biophys. Res. Commun. 231:344-347(1997).
CC -1- FUNCTION: Required for the cytoplasmic dynein-driven retrograde
CC movement of vesicles and organelles along microtubules. Dynein-
CC dynactin interaction is a key component of the mechanism of axonal
CC transport of vesicles and organelles.
CC -1- SUBUNIT: Large macromolecular complex of at least 10 components;
CC p150(glued) binds directly to microtubules and to cytoplasmic
CC dynein.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the dynactin 150 kDa subunit family.
CC -1- SIMILARITY: Contains 1 CAP-Gly domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U60312; AAB57773.1; -.
DR PIR; JCS368; JCS368.
DR MGD; MGI:107745; Dctn1.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_GLY_1.
DR PROSITE; PS00845; CAP_GLY_1; 1.
DR PROSITE; PS02045; CAP_GLY_2; 1.
KW Motor protein; Microtubule; Dynein; Coiled coil; Cytoskeleton.
FT DOMAIN 48 90 CAP-GLY.
FT DOMAIN 157 184 SER-RICH.
FT DOMAIN 214 547 COILED COIL (POTENTIAL).
FT DOMAIN 943 1049 COILED COIL (POTENTIAL).
FT DOMAIN 1185 1214 COILED COIL (POTENTIAL).
SQ SEQUENCE 1281 AA; 141721 MW; 3087FBFF0847D1EC CRC64;

Query Match 6.4%; Score 127.5; DB 1; Length 1281;
Best Local Similarity 24.5%; Pred. No. 8.8;
Matches 69; Conservative 46; Mismatches 104; Indels 63; Gaps 14;

QY 134 RPTVSASVPPYATGPPNTSYMPGMPGSGISAYPSGYPPNPSPGCPYPAGYPATTS 193
DB 151 KTRPAP-----TGAGPSSSLGPGSASAGELSSPSTPAQTPLA--APIIPTALTSP 204

QY 194 QYSPQPPVTTVGPSRDGTIS-----EDTIRASLISAVSDKLRWR-----MK 234
DB 205 --GAAPLPSPSKEEGLRAQVRLDLEKLETLRLKRSDEKAKULEKHKKHKLQLEQVQEWK 262

QY 235 EEMDGAQAEI-----NALKTEBDLKGKHKLEEM-----VTRLDQEAQV-----VD 276
DB 263 SKMQEQQADLQRLKKEARKEAKEALEAKERYMEEMADTADEIEMATLDKEMAEFRASLQ 322

QY 277 KNTIELKKDEELSSALEKMNSENNDIDEVIIPTAPLYKQILNLYABENAIEDTIFYL 336
DB 323 QEVALKERVDELTTDLLEILKAEIEKGS-----GAASSYQ--LKLEEQNA-----RL 370

QY 337 GEALRRGVLDVFLK--HV---RLLSRKQFQLEALMQRK 373
DB 371 KDALVR-MRDLSSSEKQEHVKLQKLMKKQLELVVVRQQRER 411

RESULT 14
DINA CHICK
ID DINA_CHICK STANDARD; PRT; 1224 AA.
AC P35458;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE (p150-glued).
GN DCTN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Cleveland D.W.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 172-1224 FROM N.A.
RC TISSUE=Embryonic brain;
RX MEDLINE=92098576; PubMed=1836789;
RA Gill S.R., Schroter T.A., Szilak I., Steuer E.R., Sheetz M.P.,
RA Cleveland D.W.;
RT "Dynactin, a conserved, ubiquitously expressed component of an
RT activator of vesicle motility mediated by cytoplasmic dynein."

```

```
RL J. Cell Biol. 115:1639-1650(1991).
CC -!- FUNCTION: DYNACTIN IS A MAJOR COMPONENT OF ACTIVATOR I, A 20S
CC POLYPEPTIDE COMPLEX THAT STIMULATES DYNEIN-MEDIATED VESICLE
CC TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=1;
CC Comment=At least 3 isoforms are produced;
CC Name=1;
CC IsoId=P35458-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -!- SIMILARITY: Belongs to the dynactin 150 kDa subunit family.
CC -!- SIMILARITY: Contains 1 CAP-Gly domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62773; CAA44617.2; -
CC PIR; A41642; A41642.
CC InterPro; IPR000938; CAP-Gly.
CC Pfam; PF01302; CAP_GLY; 1.
CC PROSITE; PS00845; CAP_GLY_1; 1.
CC PROSITE; PS0245; CAP_GLY_2; 1.
CC Motor protein; Microtubule; Dynein; Coiled coil; Cytoskeleton;
CC Alternative splicing.
CC CAP-GLY.
CC DOMAIN 49 91
CC FT DOMAIN 205 540
CC FT DOMAIN 936 1042
CC FT DOMAIN 1081 1117
CC FT COILED COIL (POTENTIAL).
CC FT COILED COIL (POTENTIAL).
CC SEQUENCE 1224 AA; 135562 MW; 03B7FFE68E7C01D7 CRC64;
CC
CC Query Match 6.4%; Score 127; DB 1; Length 1224;
CC Best Local Similarity 26.3%; Pred. No. 8.9;
CC Matches 76; Conservative 39; Mismatches 90; Indels 84; Gaps 16;
CC
CC QY 145 TATGPNNTSYMPGSPGISAYPSG-----YPNPSPGPGCPYPAGPYPATTS 192
CC Db 140 TARRPKPTRTPTSPSSAGTAGSGASAGSGMSSSEFSTPAQTP--LVAPVIPSPULTS 197
CC
CC QY 193 SQYSPQPPVTTVGPRDGTISDTIRASL-----ISAVSDKLRWR----- 232
CC Db 198 -----PVAPWVPS--PTKEENLRQVRDLEEKLETLKIKRNEDEKAKLEKVKIQL 248
CC
CC QY 233 -----MKEMDGAQAE-----NALKRTEDLKKGHQKLEEM-----VTRLDOEVAR 274
CC Db 249 EQVQEWKSKMQQADLQRRLEAKKEAKDALEAKERYMEEMADTADAEMATLDKEMAE 308
CC
CC QY 275 -----VDKNIELKKKDELSALEKMNQSENNDIDEVIIPTAPLYKQILNLYAEENAI 329
CC Db 309 ERAESLQGEVDSLEKKEVFLTWDLLEIKHEIEKSD-----GAASY-QVKQL-REQNA- 361
CC
CC QY 330 EDTIFYLGEALRRGVLDLDFLKL--HVRLL-----LSRKQFQLRALMQKARK 373
CC Db 362 -----RLKEALVR-MRDLSEKQEHVKLQKQMEKKNTLESILRQOREK 404
CC
CC RESULT 15
CC ID WASL HUMAN STANDARD; PRT; 505 AA.
CC AC O00401;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Neural Wiskott-Aldrich syndrome protein (N-WASP).
CC GN WASL.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97464048; PubMed=9322739;
RA Fukutaka M., Miki H., Takenawa T.;
RT "Identification of N-WASP homologs in human and rat brain.";
RL Gene 196:43-48(1997).
CC -!- FUNCTION: REGULATES ACTIN POLYMERIZATION BY STIMULATING THE ACTIN-
CC NUCLEATING ACTIVITY OF THE ACTIN-RELATED PROTEIN 2/3 (ARP2/3)
CC COMPLEX.
CC -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX; INTERACTS WITH CDC42
CC BINDS TO SH3 DOMAINS OF ASH/GRB2.
CC -!- SIMILARITY: Contains 1 CRIB domain.
CC -!- SIMILARITY: Contains 1 WH1 domain.
CC -!- SIMILARITY: Contains 2 WH2 domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D88460; BAA20128.1; -
CC Genew; HGNC:12735; WASL.
CC MIM; 605056; -
CC GO; GO:0015629; C:actin cytoskeleton; TAS.
CC GO; GO:0005083; F:small GTPase regulatory/interacting protein. .; TAS.
CC GO; GO:0008154; F:actin polymerization and/or depolymerization; TAS.
CC GO; GO:0006928; P:cell motility; TAS.
CC GO; GO:0006461; P:protein complex assembly; TAS.
CC InterPro; IPR000697; EVH1.
CC InterPro; IPR000095; PAKbox/RhoGndng.
CC InterPro; IPR001960; WH1.
CC InterPro; IPR003124; WH2.
CC Pfam; PF00786; PBD; 1.
CC Pfam; PF00568; WH1; 1.
CC Pfam; PF02205; WH2; 2.
CC SMART; SM00285; PBD; 1.
CC SMART; SM00461; WH1; 1.
CC SMART; SM00246; WH2; 2.
CC PROSITE; PS50108; CRIB; 1.
CC Actin-binding; Repeat.
CC FT DOMAIN 34 138
CC FT DOMAIN 203 216
CC FT DOMAIN 277 392
CC FT DOMAIN 405 422
CC FT DOMAIN 433 450
CC FT DOMAIN 485 505
CC FT ASP-RICH.
CC SQ SEQUENCE 505 AA; 54824 MW; BC5670A11AB63539 CRC64;
CC
CC Query Match 6.3%; Score 126.5; DB 1; Length 505;
CC Best Local Similarity 27.5%; Pred. No. 3.3;
CC Matches 56; Conservative 31; Mismatches 64; Indels 53; Gaps 12;
CC
CC QY 129 PVFSPRTVSAYPPYATGPTNTSYMPGSPGISAYPSGYPNPSGY----PGCFYPPA 184
CC Db 324 PPPSPSPSEVP-PP-----PNNRMVPPPPPPALPSSAPSGPPPPSVLGVGVAPPVPPP 377
CC
CC QY 185 GPYPATTSQYPSQPPVTVGPRDGTISDTIRASLISAVSDKLRWRKMEEMDGAQAE 244
CC Db 378 PPPP-----PPGPPPPPPGLPS-DGDHQPVT-TAGNKAALLDQIR-----EGAQ--- 418
CC
CC QY 245 NALKTEEDLK-----KGHQKLEEMVTRLDQ--EVAE-----VDKNIELLK 283
CC Db 419 --LKKVQNSRPVSCSGRDALLDQIRGQLKXSVADGQESTPTPTAPTSGIVGALMEVWQ 476
CC
CC QY 284 KKDEELSSALEKMNQSENNDIDE 307
CC Db 477 KESKAIHSS-DEDEDEDEDEDFED 499
```



```
RESULT 16
VASP_MOUSE
ID VASP_MOUSE STANDARD; PRT; 375 AA.
AC P70460; Q9R214;
DT 16-OCT-2001 (Rel. 40, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vasodilator-stimulated phosphoprotein (VASP).
GN VASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=96411679; PubMed=8812448;
RA Zimmer M., Fink T., Fischer L., Hauser W., Scherer K., Lichter P.,
RA Walter U.;
RT "Cloning of the VASP (vasodilator-stimulated phosphoprotein) genes in
RT human and mouse: structure, sequence, and chromosomal localization.";
RL Genomics 36:227-233(1996).
RN [2]
RP SEQUENCE FROM N.A., AND PHOSPHORYLATION.
RC STRAIN=C57BL/6J;
RX MEDLINE=99185054; PubMed=10085070;
RA Collins S.P., Uhler M.D.;
RT "Cyclic AMP- and cyclic GMP-dependent protein kinases differ in their
RT regulation of cyclic AMP response element-dependent gene
RT transcription.";
RL J. Biol. Chem. 274:8391-8404(1999).
CC -I- FUNCTION: Actin- and profilin-binding microfilament-associated
CC protein. May act in concert with profilin to convey signal
CC transduction to actin filament production (By similarity).
CC -I- SUBUNIT: Homotrimer (By similarity).
CC -I- SUBCELLULAR LOCATION: Focal adhesions (By similarity).
CC -I- PTM: MAJOR SUBSTRATE FOR CAMP-DEPENDENT (CAPK) AND CGMP-DEPENDENT
CC PROTEIN KINASE (CGPK) IN PLATELETS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99475; CAA67108.1; -.
DR EMBL; AF084548; AAD16045.1; -.
DR MGD; MGI:109268; Vasp.
DR InterPro; IPR000697; EVH1
KW Phosphorylation; Actin-binding.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 165 181 POLY-PRO.
FT DOMAIN 317 320 POLY-SER.
FT MOD_RES 152 152 PHOSPHORYLATION (BY PKA AND PKG) (BY
FT FT SIMILARITY).
FT MOD_RES 234 234 PHOSPHORYLATION (BY PKA AND PKG) (BY
FT FT SIMILARITY).
FT MOD_RES 273 273 PHOSPHORYLATION (BY PKA AND PKG) (BY
FT FT SIMILARITY).
FT CONFLICT 208 208 T -> A (IN REF. 2).
FT CONFLICT 287 287 MISSING (IN REF. 2).
SQ SEQUENCE 375 AA; 39693 MW; 44287D1568A3CB99 CRC64;
Query Match 6.2%; Score 123.5; DB 1; Length 375;
Best Local Similarity 20.4%; Pred. No. 3.4;
Matches 80; Conservative 56; Mismatches 121; Indels 135; Gaps 18;
QY 33 VENGSSRELVLNLTGIPVRY-RGNLYNIPICLLWLDTPYNPICFVKPTSSWTKTGK 91
DB 14 LYDSSKRWLP--AGTGPQAFSRVQIYH-----NPTANSFRVVGVR 51
```

```
QY 92 HVDANGKIYL-----PYLDWKHPRSELELELIQIMIVIFGEBPPVFSRPTV 137
DB 52 KMOPDQVWINCAIIRGVKNQATPIPHWRDAR-----QVWGLNFGSKEDAIQFATG 104
QY 138 SASY-----PYTATGPNNTSYMPGMS-----GISA YPS 167
DB 105 MANALEALGGGPP-PAPAPPAWSAQNG-PSPELEQQKQPEHMERVRVSNAGGPPAPPA 162
QY 168 GYPNPSGYPGCPYPAGYPATTSSQY-----PSQPPV-TTVGSPSRDGTISED 215
DB 163 GGPPPPGPPPPGPP--PPGGLPSSGVSGAGHAGAAPAPPLPTTQGPNSGGS-GAP 219
QY 216 TIRASLISAVSDKLRWRKSEEM-----DGAQALNA-LKRTEDLKKG 257
DB 220 GLAAATAGA---KLIRKVSQEEASGGPLAPKAENSRSTGGGLMEEMNAMLARRRKATQVG 276
QY 258 HQLEEMVTRLDQEVAEVDKNIELKKKDELSALEKWN-----QSEN 302
DB 277 EKPKDESASQSEEARLPAQSEFVRPWEKNSTTLPRMKSSSVTSEAHPTPCSSDD 336
QY 303 ND-----IDEVIITPAFLYKQILNLYAE 326
DB 337 SDLERVKQELLEVRKELQMKKEIIEVFQOE 368
```

```
RESULT 17
WASL_RAT
ID WASL_RAT STANDARD; PRT; 501 AA.
AC O08816;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neural Wiskott-Aldrich syndrome protein (N-WASP).
GN WASL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97464048; PubMed=9322739;
RA Fukuoka M., Miki H., Takenawa T.;
RT "Identification of N-WASP homologs in human and rat brain.";
RL Gene 196:43-48(1997)
CC -I- FUNCTION: REGULATES ACTIN POLYMERIZATION BY STIMULATING THE ACTIN-
CC NUCLEATING ACTIVITY OF THE ACTIN-RELATED PROTEIN 2/3 (ARP2/3)
CC COMPLEX (BY SIMILARITY).
CC -I- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX; INTERACTS WITH CDC42
CC BINDS TO SH3 DOMAINS OF ASH/GRB2 (BY SIMILARITY).
CC -I- SIMILARITY: Contains 1 CRIB domain.
CC -I- SIMILARITY: Contains 1 WH1 domain.
CC -I- SIMILARITY: Contains 2 WH2 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D88461; RAA21534.1; -.
DR InterPro; IPR000697; EVH1.
DR InterPro; IPR000095; PAKDOX/RhoGndg.
DR InterPro; IPR001960; WH1.
DR InterPro; IPR003124; WH2.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF00568; WH1; 1.
DR Pfam; PF02205; WH2; 2.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00461; WH1; 1.
DR SMART; SM00246; WH2; 2.
```


"The complete genome of the hyperthermophilic bacterium Aquifex aeolicus".
 Nature 392:353-358(1998).
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
 CC rad50/mre11 complex possesses single-strand endonuclease activity
 CC and ATP-dependent double-strand-specific exonuclease activity.
 CC Rad50 provides an ATP-dependent control of mre11 by unwinding
 CC and/or repositioning DNA ends into the mre11 active site (By
 CC similarity).
 CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
 CC -!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000718; AAC07092.1; -;
 CC PIR; A70387; A70387.
 CC HAMAP; MF_00449; -; 1.
 CC InterPro; IPR003439; ABC transporter.
 CC InterPro; IPR007523; DUF498.
 CC InterPro; IPR004592; SbcC.
 CC InterPro; IPR002017; Spectrin.
 CC Pfam; PF04430; DUF498; 1.
 CC TIGRFAMs; TIGR00618; sbcc; 1.
 CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
 CC NP BIND 32 39 ATP (BY SIMILARITY).
 CC FT DOMAIN 160 826 COILED COIL (POTENTIAL).
 CC SEQUENCE 978 AA; 115897 MW; 9B0F2BF51ADD1151 CRC64;
 CC
 CC Query Match 6.1%; Score 122; DB 1; Length 978;
 CC Best Local Similarity 29.1%; Pred. No. 13;
 CC Matches 43; Conservative 34; Mismatches 55; Indels 16; Gaps 6;
 CC
 CC 233 MKEMDGAQELNALKTEEDLKGKQKLEMTVRLD--QVAVVDKNI-----ELIKKKD 286
 CC 307 LKDELSFAQELNKNIEAKFKREKEKELEHRLKLOEIKELQSLSSLSKE 366
 CC 287 EELSSALEKMESENNDI-DEVIPTAPLYKQILNLYABENALEDTFYVIGEARRGVI 345
 CC 367 REVEQAQEFEDLSERVEKGLVAETEKLEIKELFSEE---EYSLMKERL---LV 420
 CC 346 DLDFVLKHVLLSEKQPOLRALMQARK 373
 CC 421 ELQKKLKE--LKEKEGQLENLTQKYKE 445
 CC
 CC RESULT 20
 CC DAM2 MOUSE
 CC ID DAM2_MOUSE STANDARD; PRT; 1068 AA.
 CC AC Q80U19; Q810J5;
 CC DT 10-OCT-2003 (Rel. 42, Created)
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Disheveled associated activator of morphogenesis 2.
 CC GN DAM2 OR KIAA0381.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC OX NCBI_TaxID=10090;
 CC RN [1]
 CC RP SEQUENCE FROM N.A. (ISOFORM 1).
 CC RC TISSUE=Brain;
 CC RX MEDLINE=22579291; PubMed=12693553;
 CC RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 CC Nakajima D., Nagase T., Ohara O., Koga H.;
 CC "Prediction of the coding sequences of mouse homologues of KIAA gene:
 CC II. The complete nucleotide sequences of 400 mouse KIAA-homologous
 CC cDNAs identified by screening of terminal sequences of cDNA clones

randomly sampled from size-fractionated libraries.";
 DNA Res. 10:35-48(2003).
 [2]
 RN SEQUENCE FROM N.A. (ISOFORM 2).
 RP TISSUE=Pancreas;
 RC MEDLINE=22388257; PubMed=12477932;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ussin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 CC and mouse cDNA sequences";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q80U19-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q80U19-2; Sequence=VSP_008005, VSP_008006;
 CC Note=No experimental confirmation available;
 CC -!- SIMILARITY: Belongs to the formin homology family.
 CC -!- SIMILARITY: Contains 1 Formin homology 1 (FH1) domain.
 CC -!- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to what seems
 CC to be the presence of intronic sequence in the cDNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AK122266; BAC65548.1; ALT_SEQ.
 CC EMBL; BC050043; AAH50043.1; -;
 CC MGD; MGI:1923691; Daam2.
 CC InterPro; IPR003104; FH2.
 CC Pfam; PF02181; FH2; 1.
 CC SMART; SM00498; FH2; 1.
 CC Coiled coil; Alternative splicing.
 CC KW DOMAIN 433 516 COILED COIL (POTENTIAL).
 CC FT DOMAIN 518 604 FH1 (PRO-RICH).
 CC FT DOMAIN 595 1024 FH2.
 CC FT DOMAIN 685 721 COILED COIL (POTENTIAL).
 CC FT DOMAIN 862 1012 COILED COIL (POTENTIAL).
 CC FT DOMAIN 522 525 POLY-PRO.
 CC FT DOMAIN 541 548 POLY-PRO.
 CC FT DOMAIN 553 562 POLY-PRO.
 CC FT VARSPLIC 516 524 TGPVSSPPP -> VRGHHHPPI (in isoform 2).
 CC FT /FTID=VSP_008005.
 CC FT VARSPLIC 525 1068 Missing (in isoform 2).
 CC FT /FTID=VSP_008006.
 CC SQ SEQUENCE 1068 AA; 123534 MW; 6C5E16BE566B932 CRC64;
 CC
 CC Query Match 6.1%; Score 121.5; DB 1; Length 1068;
 CC Best Local Similarity 23.7%; Pred. No. 15;
 CC Matches 65; Conservative 45; Mismatches 99; Indels 65; Gaps 13;
 CC
 CC 145 TATGPNTSYWPGMPSGISAYPSGY----PPNPSTGYGCFYPAGYPATTSQPSQPP 200

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences."
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Binds calmodulin in a calcium dependent manner. May
CC function as scaffolding or signaling protein.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-bound (By
CC similarity).
CC -!- MISCELLANEOUS: The name 'Zinedin' probably originates from the
CC name of the famous soccer player from Marseille (Zinedine
CC Zidane).
CC -!- SIMILARITY: Belongs to the WD-repeat striatin family.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF212940; AAP29527.1; -
CC EMBL; BC004910; AAO04910.1; -
CC Genew; HGNC:15721; STRN4.
CC GO; GO:0005737; C:cytoplasm; TAS.
CC GO; GO:0005624; C:membrane fraction; TAS.
CC GO; GO:0005116; F:calmodulin binding; TAS.
CC GO; GO:0005198; F:structural molecule activity; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
CC GO; GO:0007268; P:synaptic transmission; TAS.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 5.
CC ProDom; PD000018; WD40; 2.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00678; WD REPEATS 1; 1.
CC PROSITE; PS50082; WD REPEATS 2; 4.
CC PROSITE; PS50294; WD REPEATS REGION; 1.
CC Calmodulin-binding; Repeat; WD repeat; Coiled coil.
CC DOMAIN 69 136 COILED COIL (POTENTIAL).
CC DOMAIN 165 182 CALMODULIN-BINDING (POTENTIAL).
CC REPEAT 436 475 WD 1.
CC REPEAT 489 528 WD 2.
CC REPEAT 542 581 WD 3.
CC REPEAT 587 628 WD 4.
CC REPEAT 635 674 WD 5.
CC REPEAT 677 716 WD 6.
CC REPEAT 723 752 WD 7.
CC SITE 71 79 CAVEOLIN-BINDING (POTENTIAL).
CC DOMAIN 6 14 POLY-ALA.
CC CONFLICT 402 404 LAD -> GTR (IN REF. 2).
CC SEQUENCE 753 AA; 80581 MW; 4DA016A8FF7EDB5E CRC64;

Query Match 6.0%; Score 119.5; DB 1; Length 753;
Best Local Similarity 24.8%; Pred. No. 13;
Matches 66; Conservative 33; Mismatches 92; Indels 75; Gaps 13;

QY 138 SASYPPTATGTPNTSYMPGSPGSIASYPGYPNPISGY-----PGCPYPAGYPATTSS 193
Db 15 SSCRPLGSGAGFOPTQAAP-----VSAPAPG--PGFAGKGGGGSGPGTAGPEPLSL-- 65

QY 194 QVPSQPPTVTGSPRDTISDTIRASLISAVSDKLRWRKMKEMDGAQAEALAKRTEED 253
Db 66 -----PGILHFIQHEWARFE-----AEKARW-----EAEAEALQAQVAFLOQ 102

QY 254 LKKGHOKLE-EMVTR-----LDQFAEVDKNIKLLKKDELSALEK--NENQSENN 303
Db 103 ERKQGNLKTDLVRRIRKIMLEYALKQERAKYHK-----LKFQTDNLQGEKKAADVSEQVNG 157

QY 304 DIDEVIPTAPLY-----KQILNLYAEENAIETDIFVLGEALRGVIDLDFVLKHKVLLSR 359
Db 158 PVESVLENSPLWKEGRQLLQRYLEEVGYTDI-----LDMRSKRVSLIG 204

QY 360 KQFQIRALMQKARKT-----AGLS 378
Db 205 RSLLENGAVEPSEGAPRAPPGPAGLS 230

RESULT 23
MK07_HUMAN
ID MK07_HUMAN STANDARD; PRT; 815 AA.
AC Q13164; Q16634;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mitogen-activated protein kinase 7 (EC 2.7.1.37) (Extracellular
DE signal-regulated kinase 5) (ERK-5) (ERK4) (BMK1 kinase).
GN MAPK7 OR PRKM7 OR ERK5 OR ERK4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95279403; PubMed=7759517;
RA Zhou G., Bao Z.Q., Dixon J.E.;
RT "Components of a new human protein kinase signal transduction
RT pathway."
RL J. Biol. Chem. 270:12665-12669(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95374539; PubMed=7646528;
RA Lee J.-D., Ulevitch R.J., Han J.;
RT "Primary structure of BMK1: a new mammalian map kinase."
RT Biochem. Biophys. Res. Commun. 213:715-724(1995).
CC -!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND
CC NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- ENZYME REGULATION: Activated by tyrosine and threonine
CC phosphorylation (By similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN
CC HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE
CC IN LIVER.
CC -!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN
CC TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.
CC -!- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN
CC THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
CC ROLE, IS ABSENT.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. MAP
CC kinase subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U25278; AAA81381.1; -
CC EMBL; U29725; AAA82931.1; -
CC EMBL; U29726; AAA82932.1; -
CC EMBL; U29727; AAA82933.1; -
CC FIR; B56708; B56708.
CC HSSP; P24941; IHCL.
CC Genew; HGNC:6880; MAPK7.
CC MIM; 602521; -
CC GO; GO:0004707; F:MAP kinase activity; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
CC InterPro; IPR003527; MAP_kin.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR008271; Ser_thr_pkin_AS.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 17.0223 Seconds
(without alignments)
2152.993 Million cell updates/sec

Title: US-09-804-690-2

Perfect score: 1996
Sequence: 1 MMSKYYRDLTVRQTVNVIA.....FQLRALMQARKTAGLSLDLY 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_78.*

1: _pir1.*
2: _pir2.*
3: _pir3.*
4: _pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	627.5	31.4	404	2 H88650	protein C09G12.9 [
2	396	19.8	83	2 I48283	gene CC2 protein -
3	151	7.6	169	2 T34520	hypothetical prote
4	142	7.1	212	2 S74288	hypothetical prote
5	141.5	7.1	485	2 T37550	hypothetical colle
6	137.5	6.9	503	1 LURB11	annexin XI - rabbi
7	137	6.9	488	1 LUHU7	annexin VII, long
8	135.5	6.8	463	2 S29170	annexin VII - mous
9	135.5	6.8	827	2 T39608	zinc finger transc
10	133	6.7	198	2 D70509	hypothetical prote
11	133	6.7	505	2 S72273	actin-depolymerizi
12	130	6.5	1006	2 T42731	atrophin-1 related
13	129	6.5	669	2 T28754	hypothetical prote
14	128	6.4	505	2 A53152	annexin XI - human
15	127.5	6.4	1281	2 JCS368	dynactin 1 - mouse
16	127	6.4	279	2 T05421	hypothetical prote
17	124	6.2	553	2 G83385	hypothetical prote
18	123.5	6.2	1181	2 C86349	F8K7.4 protein - A
19	122.5	6.1	139	2 H84809	hypothetical prote
20	122.5	6.1	503	1 LUBO11	annexin XI form A
21	122.5	6.1	505	1 S23447	annexin XI form B
22	122.5	6.1	1902	2 C97702	cell surface antig
23	122	6.1	792	2 T49989	hypothetical prote
24	122	6.1	978	2 A70387	conserved hypothet
25	121.5	6.1	214	2 T10737	extensin-like cell
26	121.5	6.1	214	2 T09854	proline-rich celly
27	121	6.1	887	1 S57219	1-phosphatidylinos
28	120.5	6.0	338	2 I53043	transforming prote
29	120.5	6.0	729	2 E70803	hypothetical prote

30	120	6.0	678	2 H88187	protein C18H9.8 [i
31	120	6.0	1605	2 T31435	DNA-directed RNA p
32	119.5	6.0	815	2 B56708	extracellular sign
33	119.5	6.0	1189	2 T42726	guanine nucleotide
34	119.5	6.0	1613	2 S39059	protein BRG1 - hum
35	119.5	6.0	1733	1 RNEY2L	DNA-directed RNA p
36	119.5	6.0	1802	2 H88444	protein C26E6.12 [
37	119	6.0	380	2 S51797	vasodilator-stimul
38	118.5	5.9	384	2 S51796	vasodilator-stimul
39	118.5	5.9	1647	2 S45252	SNF2beta protein -
40	118	5.9	929	2 C96623	hypothetical prote
41	117.5	5.9	262	2 A54889	IgE-binding protei
42	117.5	5.9	1184	2 G01763	atrophin-1 - human
43	117	5.9	437	2 T14192	extensin homolog T
44	117	5.9	609	2 S46019	YSW1 protein - yea
45	117	5.9	964	2 T21865	hypothetical prote
46	117	5.9	1453	2 S21626	collagen alpha 1(I
47	117	5.9	1752	1 S26849	DNA-directed RNA p
48	116.5	5.8	530	2 T48627	hypothetical prote
49	116.5	5.8	938	2 A56731	chromatin assembly
50	116.5	5.8	1024	2 T27631	hypothetical prote
51	116.5	5.8	1030	2 H88859	protein ZC518.2 [i
52	116.5	5.8	1896	1 RNFF2L	DNA-directed RNA p
53	116	5.8	338	1 TWMSPB	transforming prote
54	116	5.8	554	2 F86244	hypothetical prote
55	116	5.8	809	2 S16266	cellulose synthase
56	116	5.8	1046	2 T42734	cytoplasmic linker
57	116	5.8	1633	2 JC5056	polybromo 1 - chic
58	116	5.8	2715	2 T13049	eyelid - fruit fly
59	115.5	5.8	1790	2 S67593	transport protein
60	115	5.8	605	1 QBE3R	BVPF2 (EC-RF3) pro
61	115	5.8	751	2 S68957	adhesive plaque pr
62	114.5	5.7	574	2 T43556	Wiskott-Aldrich sy
63	114.5	5.7	574	2 T38819	wiskott-aldrich sy
64	114.5	5.7	622	2 I48733	protein kinase rok
65	114.5	5.7	853	2 T51505	hypothetical prote
66	114.5	5.7	1422	2 T24212	hypothetical prote
67	114	5.7	176	2 A86441	hypothetical prote
68	114	5.7	228	2 S35504	extensin-like prot
69	114	5.7	242	2 S35060	tropomyosin - hydr
70	114	5.7	342	2 A24263	myosin heavy chain
71	114	5.7	361	2 AF3370	hypothetical prote
72	114	5.7	791	2 S67265	hypothetical prote
73	114	5.7	821	2 S67087	hypothetical prote
74	113.5	5.7	272	2 T25608	hypothetical prote
75	113.5	5.7	469	2 S44655	ZK353.8 protein -
76	113.5	5.7	839	2 T04859	extensin homolog F
77	113	5.7	902	2 A56823	DNA-directed RNA p
78	113	5.7	1053	2 A16642	dynactin - chicken
79	112.5	5.6	451	2 B70792	hypothetical prote
80	112.5	5.6	1008	2 T04462	hypothetical prote
81	112.5	5.6	1069	2 D85383	hypothetical prote
82	112.5	5.6	1133	2 T22976	hypothetical prote
83	112.5	5.6	1548	2 T25808	hypothetical prote
84	112.5	5.6	1940	1 A24922	myosin heavy chain
85	112	5.6	504	2 A49467	occludin - chicken
86	112	5.6	562	2 S75308	DNA ligase (EC 6.5
87	112	5.6	671	2 T36037	probable export as
88	112	5.6	1046	2 T42720	cytoplasmic linker
89	112	5.6	1179	2 F71190	probable chromosom
90	112	5.6	1859	1 A34092	DNA-directed RNA p
91	112	5.6	1862	2 T29959	DNA-directed RNA p
92	111.5	5.6	177	2 E70318	glycine/proline-ri
93	111.5	5.6	400	2 E70318	hypothetical prote
94	111.5	5.6	491	2 S14182	DNA-directed RNA p
95	111.5	5.6	513	2 T03916	hypothetical prote
96	111.5	5.6	650	2 S14181	DNA-directed RNA p
97	111.5	5.6	834	2 S44866	R0SD3.4 protein -
98	111.5	5.6	884	2 D96730	unknown protein F5
99	111.5	5.6	924	2 T06636	hypothetical prote
100	111.5	5.6	1300	2 T18364	ro-3 protein - Neu

I48283
 gene CC2 protein - mouse (fragment)
 C;Species: Mus musculus (house mouse)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
 C;Accession: I48283
 R;Maucuer, A.; Camonis, J.H.; Sobel, A.
 Proc. Natl. Acad. Sci. U.S.A. 92, 3100-3104, 1995
 A;Title: Statmin interaction with a putative kinase and coiled-coil-forming protein domain
 A;Reference number: I48282; MUID:95241452; PMID:7724523
 A;Accession: I48283
 A;Status: preliminary; translated from GB/EMBL/DBDB
 A;Molecule type: mRNA
 A;Residues: 1-83 <RES>
 A;Cross-references: EMBL:X82319; NID:g791075; PIDN:CAA57762.1; PID:g791076
 C;Genetics:

A;Gene: SGD:STP22

Db 315 EFKXTEAIRSDTSGHFQRLILSLSCQNRDESNVDMSLVQRDVQELIYAAGE----NRL 370

Qy 346 DLDVFLKHVRLSRKQFQLRALMQKARKTAG 376

Db 371 GTDESKFNVLCSRSRAHLVAVFNEYQRM TG 401

RESULT 7

LUHUV

annexin VII, long form - human

N;Alternate names: synexin

N;Contains: annexin VII, long form; annexin VII, short form

C;Species: Homo sapiens (man)

C;Date: 30-Sep-1992 #sequence revision 26-Jan-1996 #text_change 22-Jun-1999

C;Accession: A54467; A32554; A39513; B39513

R;Shirvan, A.; Srivastava, M.; Wang, M.G.; Cultraro, C.; Magendzo, K.; McBride, O.W.; Pol Biochemistry 33, 6888-6901, 1994

A;Title: Divergent structure of the human synexin (annexin VII) gene and assignment to c A;Reference number: A54467; MUID:94264005; PMID:7515686

A;Accession: A54467

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-488 <SHI>

R;Burns, A.L.; Magendzo, K.; Shirvan, A.; Srivastava, M.; Rojas, E.; Alijani, M.R.; Poll Proc. Natl. Acad. Sci. U.S.A. 86, 3798-3802, 1989

A;Title: Calcium channel activity of purified human synexin and structure of the human s A;Reference number: A32554; MUID:99264510; PMID:2542947

A;Accession: A32554

A;Molecule type: mRNA

A;Residues: 1-145,168-488 <BUR>

A;Cross-references: EMBL:J04543; NID:G338243; PIDD:AAA36616.1; PTD:G338244

R;Magendzo, K.; Shirvan, A.; Cultraro, C.; Srivastava, M.; Pollard, H.B.; Burns, A.L. J. Biol. Chem. 266, 3228-3232, 1991

A;Title: Alternative splicing of human synexin mRNA in brain, cardiac, and skeletal musc A;Reference number: A39513; MUID:91131630; PMID:1825209

A;Accession: A39513

A;Molecule type: mRNA

A;Residues: 137-145,168-176 <MAG>

A;Cross-references: EMBL:J05732

A;Accession: B39513

A;Molecule type: mRNA

A;Residues: 137-176 <MA2>

A;Comment: Annexins undergo reversible, calcium-dependent binding to membrane phospholip derstood.

C;Comment: The long form of annexin VII is more prevalent in brain, heart, and skeletal C;Genetics:

A;Gene: GDB:ANX7

A;Cross-references: GDB:369042; OMIM:186360

A;Map position: 10q21.1-10q21.2

A;Introns: 17/3; 87/1; 124/1; 145/3; 167/3; 202/2; 233/3; 271/3; 328/3; 385/3; 411/1; 44 C;Superfamily: annexin VII; annexin repeat homology

C;Keywords: alternative splicing; calcium binding; calcium channel; duplication; endonex F;1-488/Product: annexin VII, long form #status predicted <MA1>

F;1-145,168-488/Product: annexin VII, short form #status predicted <MA3>

F;188-259/Domain: annexin repeat homology <AX1>

F;199-215/Region: endonexin fold #status predicted

F;260-331/Domain: annexin repeat homology <AX2>

F;271-287/Region: endonexin fold #status predicted

F;343-415/Domain: annexin repeat homology <AX3>

F;355-371/Region: endonexin fold #status predicted

F;419-488/Domain: annexin repeat homology <AX4>

F;430-446/Region: endonexin fold #status predicted

Query Match 6.9%; Score 137; DB 1; Length 488;

Best Local Similarity 20.7%; Pred. No. 0.59; Gaps 17;

Matches 85; Conservative 34; Mismatches 92; Indels 200; Gaps 17;

Qy 140 SYPPYTATG-PPNTSYMPG-----MPSG:SAVPSGYP----- 170

Db 2 SYEGPTGYPPFPFPAGQESSFFPSGQVYPSGFPFMGGGAYPQVPSGYPGAGY 61

Qy 171 -----PNPSGYPGCPYPAG-YPATSSQ-----YPSQPPVTYV----- 205

Db 62 APGGYPAPGGYPGAPQPGGAPSYGVPVPGQGVPPGAGFSGYP-QPPSQSYGGGAQV 120

Qy 206 -----PSRDT-----ISEDTIR 218

Db 121 PLPGGFGGMPGQYPGGQPTYPISQINTDSFSSYVPFSPVSLDYSSEPATVTVQTGTIR 180

Qy 219 -ASLISAVSK--LRWPKSEMDGAQALNALKTEED----- 253

Db 181 PAANFDIAIRDAEILRKAMKGFQTEQAIIVDVVANRSNDQOKIKAAFKTSYGGKDLIKDLK 240

Qy 254 -----LKGHQ-----KLEEMVTRLDQVAVV----- 275

Db 241 SELSGNMEELILALFMPPTTYDAWSLRKAMQAGTQERVLIEILCTRNQEIIRIVRCYQ 300

Qy 276 -DKNIELKKKDELSALEKM-----ENQSENNDIDEVIIPTADLYKQILNLYA 324

Db 301 SEFGRLDEKDIRDTSGHFRLVSMQGNRDNQSNHQM-----A 342

Qy 325 EENAIETITFVLGEALRRGVLDLVFLKHVRLSRKQFQLRALMQKARKTA 375

Db 343 QEDA--QLIYQAGE---GRLGTDESCFNILATRSFPQLRAIWEYSRMA 387

RESULT 8

S29170

annexin VII - mouse

N;Alternate names: synexin

C;Species: Mus musculus (house mouse)

C;Date: 25-Feb-1994 #sequence revision 01-Sep-1995 #text_change 13-Aug-1999

C;Accession: S29170; S46209; S51173

R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B. Biochem. J. 289, 735-741, 1993

A;Title: Mouse synexin (annexin VII) polymorphisms and a phylogenetic comparison with o A;Reference number: S29170; MUID:93168121; PMID:7916616

A;Accession: S29170

A;Molecule type: mRNA

A;Residues: 1-463 <ZHA>

A;Cross-references: EMBL:L13129

R;Zhang-Keck, Z.Y.; Srivastava, M.; Kozak, C.A.; Caohuy, H.; Shirvan, A.; Burns, A.L.; Biochem. J. 301, 835-845, 1994

A;Title: Genomic organization and chromosomal localization of the mouse synexin gene. A;Reference number: S46209; MUID:94330961; PMID:8053909

A;Accession: S46209

A;Molecule type: DNA

A;Residues: 1-463 <ZHF>

R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B. submitted to the EMBL Data Library, May 1993

A;Reference number: S51173

A;Accession: S51173

A;Molecule type: mRNA

A;Residues: 1-144, 'S', 146-303, 'A', 305-463 <ZHW>

A;Cross-references: EMBL:L13129; NID:G293293; PIDD:AAA37238.1; PTD:G293294

C;Genetics:

A;Gene: MGI:Anx7

A;Cross-references: MGI:88031

A;Map position: 14

A;Introns: 19/3; 87/1; 121/1; 142/3; 177/1; 208/3; 246/3; 303/3; 360/3; 366/1; 423/3 C;Superfamily: annexin VII; annexin repeat homology

C;Keywords: calcium channel; ion channel

F;163-234/Domain: annexin repeat homology <AX1>

F;235-306/Domain: annexin repeat homology <AX2>

F;318-390/Domain: annexin repeat homology <AX3>

F;394-463/Domain: annexin repeat homology <AX4>

Query Match 6.8%; Score 135.5; DB 2; Length 463;

Best Local Similarity 21.8%; Pred. No. 0.68; Gaps 17;

Matches 78; Conservative 42; Mismatches 105; Indels 133; Gaps 17;

Qy 126 GEEPT-VFVSRTVSAVPPYATGTPNTSYMPG-----MPSGISAYPSGY- 169

Db 30 GQYPSGFPFMGGGAYPAPGPGYPCAGGYPAGGYPALSGPGGALSGPPAYPGQG 89

Qy 170 ---PPNPSGYPGCPYPPA-----GP-----YPATSSQYPSQPPVTTV 204

Db 90 FGAPPGAGFGYPQPPAQSYGGGPAQVVPVGGFFGGQMPGQYPC-GQAPYPSQAAMTQ 148
QY 205 GPS-----RDGTISBETIRA-----SLISAVSDK----- 228
Db 149 GTQGTILPASNFDMRDAEILRKAMKMGFTDQAIIVDVVSNRSDORQOIKAAFKTMYGK 208
QY 229 -LRWAKMEMDGAQAEALNA---LKRTEED---LKKGHQ-----KLEEMVTRLDQEVA 273
Db 209 DLIKDLKSELSGNMEEILALEMPSFTYDAWSLRKAMQAGAQOERVLLEILCTRINQEUR 268
QY 274 EVDKNIELLKKD-----EELSSALEKM-----ENQSENNDIDIVIPTAPLYK 317
Db 269 DIVRCYQLEFGRDLKEDIRSDTSGHFERLLVSMQCNRDEROSVNHQM----- 316
QY 318 QILNLIYAENATEDITFYLGEALRGVIDLDVFLKXHVLLSRKQFOLRALMOKARKTA 375
Db 317 -----AQEDA--QRLYQAGE---GRUGTDESCFNMLIATRSFOLKATMEAYSEMA 362
RESULT 9
T39608
zinc finger transcription factor - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: T39608
R:lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21866
A:Accession: T39608
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-827 <LYN>
A:Cross-references: EMBL:AL023554; PIDN:CAAL19035.1; GSPDB:GNC0067; SPDB:SPBC16G5.16
A:Experimental source: strain 972h; cosmid c16G5
C:Genetics:
A:Map position: 2
A:Introns: 30/3; 41/1
C:Superfamily: GAL4 zinc binuclear cluster homology
F:11-47/Domain: GAL4 zinc binuclear cluster homology <GL4>
Query Match 6.8%; Score 135.5; DB 2; Length 827;
Best Local Similarity 25.7%; Pred. No. 1.4;
Matches 63; Conservative 28; Mismatches 97; Indels 57; Gaps 14;
QY 7 YRDLTVROTVNVIAMYKDLKPLVLDVSVFNDGSSREL---VNLGTGTPVVRGNINYP-- 61
Db 588 YRESTYISTI-----WEAKNLLIIADNMSSGTENLDATPDVTGQLPNNFSORTSNIPRE 642
QY 62 ---ICLLMLDT-YP--YNP-----PICFVKPTSSMTIKTKGHVDANGKIYILPYLHDWKH 109
Db 643 FPOAQIFYSADAPYPGYINPAQFNAPTNPMPITYGGRTQDSQVPRQNG--YPSYSDGNVY 700
QY 110 PRSELLELIQIMIVFGEPPFVSRTVSASYPPTATGPP---NTSYMP-----GMP5 160
Db 701 PHDRVM-----INVS-----SMTANGFYVPNTYSPVFPYNTSPYFYSPTSNTMPO 748
QY 161 GISA-----YPSGYPPNPSGPGCPYPAG---PYPATSSQYPSQPPVTVTGPSRDGTIS 213
Db 749 AFQAYSQYQYQHPFPLSEQMLPLTISGVMMAPGAASKGMPYFFIQP-----PSMTNQVA 803
QY 214 EDTIR 218
Db 804 YPTVR 808
RESULT 10
D70509
hypothetical protein Rv1233c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70509

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome;
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70509
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-198 <COL>
A:Cross-references: GB:Z98260; GB:AL123456; NID:g3261826; PIDN:CA810930.1; PID:el299942;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1233c
Query Match 6.7%; Score 133; DB 2; Length 198;
Best Local Similarity 31.6%; Pred. No. 0.33;
Matches 50; Conservative 20; Mismatches 54; Indels 34; Gaps 10;
QY 126 GEEPPVFSRP-----TVSASYP--YTATGPNTSYMPGMPGSI----- 162
Db 17 GGGPPVGERPPBPQIADAPWAPPASSPMANHPAPYPPSGYP-PAYQPGYPTGYPPMPBP 75
QY 163 SAY-PSGYPP---NPSGYGCPYPP-AGPYPATTSQYPSQPPVTVGSRDGTISEDIT 217
Db 76 GGYAPFGYPPPTSGAGYGDIFYPMPFPYGGSPGYPEPGYLDGYGSPGMMTALV 135
QY 218 RASLISAVSDKL--RWRMKEMDGAQAEALNALKRTKEED 253
Db 136 --SLISALVGVLCCIGSIVGVFGAIA-INLIKQTREE 170
RESULT 11
S72273
actin-depolymerizing protein N-WASP, brain - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 23-Apr-1998 #sequence_revision 01-May-1998 #text_change 05-Nov-1999
C:Accession: S72273
R:Miki, H.; Miura, K.; Takenawa, T.
EMBO J. 15, 5326-5335, 1996
A:Title: N-WASP, a novel actin-depolymerizing protein, regulates the cortical cytoskeleton
A:Reference number: S72273; MUID:97050838; PMID:8895577
A:Accession: S72273
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-505 <WIK>
A:Cross-references: EMBL:D67066; NID:g1644231; PIDN:BAAL1082.1; PID:dl011742; PID:g16442;
A:Experimental source: brain
Query Match 6.7%; Score 133; DB 2; Length 505;
Best Local Similarity 27.0%; Pred. No. 1.1;
Matches 58; Conservative 31; Mismatches 66; Indels 60; Gaps 12;
QY 126 GEEPPVFSR-PTVSASYPPTATG-----PNTSYMPGMPGSIAYPSGYPPNP----- 173
Db 308 GAPPFPFSGRAPTAAPPPPPPSRPGVGAAPPFPPPPPPPLPALPSAFSPGPPPPPPUSV 367
QY 174 SGYPGCPVPPAGYPATTSQYPSQPPVTVTGPSRDGTISEDITRASLISAVSDKLRWRM 233
Db 368 SGSVAPPFPPPPPPPP-----PGPPP--PPGLPSGDGHQVFTPAGSK-AALLDQIR--- 414
QY 234 KEEMDGAQAEALNALKRTKEEDLK---KGHQKLEEMVTRLDQ--EVAEVDKN----- 278
Db 415 ---EGAQ-----LKKVEQNSRPVSCGRDALLDQIRQIGLQKSVTDAPESPAPAPTS 465
QY 279 -----IELKKOEELSSALEKMNQSENNDIDE 307
Db 466 GIVGALMEYMQKRSKAIHSS-----DEDEDEDDDE 495
RESULT 12
T42731

```
atrophin-1 related protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T42731
R;Khan, F.A.; Margolis, R.L.; Loev, S.I.; Sharp, A.H.; Li, S.H.; Ross, C.A.
submitted to the EMBL Data Library, December 1995
A;Description: cDNA sequence and expression of an atrophin-1 (DRELA disease gene) relate
A;Reference number: 222250
A;Accession: T42731
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1006 <KHA>
A;Cross-references: EMBL:U44091; NID:g1297310; PID:g1209103; PIDN:AAA98970.1
C;Genetics:
A;Gene: ARP

Query Match          6.5%; Score 130; DB 2; Length 1006;
Best Local Similarity 20.0%; Pred. No. 3.8;
Matches 64; Conservative 41; Mismatches 111; Indels 104; Gaps 11;

QY 44 NLTCGTPVRYRGNINYPICLWLLDTYFYNPPICFVKPTSSMTIKTKGHVDANGKIYLPY 103
Db 388 HLSGPFSPSMANL-----PPPALKPLSLST----- 415
QY 104 LHDWKHPRSELLELIQIMIVIFGEEPPVFSRP-----TVSASYPPTYATGP----- 149
Db 416 ---HHPPSAHPPPLQLM-----PQSQPLPSSPAQPPGLTQSLSLPPAAASHPTTGLHQV 467
QY 150 -----NTSYMFGMPGSGI---SAYPSGYPP-----NPSGYPGCPY 181
Db 468 FQSOPFPQHPFVPGPPPIPTPPSCPTSTTPAGPSSSSQPPCSAAVSSGGNVPAGPSCPL 527
QY 182 PPAG-PYPATNTSSQYSPQPPVTVVTPGSRDGTISDTIRASLIISAVSKLRWKEEMDVG- 239
Db 528 PAVQIKKEALDEABEPPSPPPPPSPPEPTVVDTPSHASQSA-----RFYKHLDRG 579
QY 240 ----AQAE-----NALKRTREEDLKKGHKLEEMVTRLDQEVAEVDKNIELLKKKDE 287
Db 580 YNSCARDLYFMPLAGSKLAKKREAEIEKAKREAEQAKREAREKEKEKEKEKEKEKEKE 639
QY 288 ELSALEKMNQSENNDIDE 307
Db 640 EAERAAQKASSSAHEGRISD 659

RESULT 13
T28754
hypothetical protein T08B2.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28754
R;Blanchard, M.; Wamsley, P.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid T08B2.
A;Reference number: 220518
A;Accession: T28754
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-669 <HLA>
A;Cross-references: EMBL:AF000263; PIDN:AAC48191.1; GSPDB:GN00019; CESP:T08B2.5
A;Experimental source: strain Bristol N2; clone T08B2
C;Genetics:
A;Gene: CESP:T08B2.5
A;Map position: 1
A;Introns: 33/3; 68/3; 108/1; 177/1; 216/1; 242/3; 297/1; 418/1; 511/3; 623/1

Query Match          6.5%; Score 129; DB 2; Length 669;
Best Local Similarity 20.7%; Pred. No. 2.6;
Matches 81; Conservative 60; Mismatches 193; Indels 58; Gaps 14;

QY 39 SRELNVLTGTPVRYRGNINYPICLWLLDTYFYNPPICFVKPTSSMTIKTKGHVDAN-- 96
Db 158 AKMLLNTYKSPIKDKDVQSWCRDSMSKLIQQQMTLTAGPGKNSVATGSLAQGNMT 217

QY 97 -GKIYLPYLHDWKHPRSELLELIQIMIVIFGEE-----PPVFSRPTVSASY-PPYATATGPP 150
Db 218 GAETAAALSKANAVRQASQGGQIMTGLPAPAAANMIPPNFSPVPPNLSVFPFPMQATQPE 277
QY 151 NTSYMPGM---PSGISAYPSGYPPNPSGYPCP-----YPPA-GPYPATNTSS 193
Db 278 HQNGVIGTQTPKGL--LPRYLPNPLTFAHDNLYGVVDPITKFCYDSATGYFNNATS 335
QY 194 Q-----YPSQPP--VTVVTPGSRDGTISDTIRASLIISAVSKLRWKEEMDGA 240
Db 336 QWCNWLTHRTYFPVETPAVINSADPEERKMKNEEGPKTAQDLVKDMAKWAKQEKDKK 395
QY 241 QAEINALKRTEEDLKKGHKLEEMVTRLDQEVAEVDKNIELLKKKDELSALEKMNQ 300
Db 396 KVQISIKGKETGIELKNVFSNEKQKRIQAALFDDD-----EEDQEEVEVRGRSRSEPS 451
QY 301 ENNDIDEVIITAPLYKQILNLY-AEENAIJEDTIFYLGEALRRGVIDLDFLKHVRL--L 357
Db 452 TSSSFSV--PLAPRKSTLQDVRDAMERALLYDETKKTCMLCKRAFSDEVLKRVHXSDDL 509
QY 358 SRKQFQRL-----ALMQARKTAGLSDL 380
Db 510 HRNLEAKRAEWGRETAAKLQEEEDASAPDL 541

RESULT 14
A53152
annexin XI - human
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
C;Accession: A53152
R;Misaki, Y.; Pruijn, G.J.M.; van der Kemp, A.W.C.M.; van Venrooij, W.J.
J. Biol. Chem. 269, 4240-4246, 1994
A;Title: The 56K autoantigen is identical to human annexin XI.
A;Reference number: A53152; MUID:94140847; PMID:7508441
A;Accession: A53152
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-505 <MIS>
A;Cross-references: GB:L19605; NID:g457128; PIDN:AAA19734.1; PID:g457129
C;Genetics:
A;Gene: GDB:ANX11
A;Cross-references: GDB:313076
A;Map position: 9q11-9q22
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid bi
F;203-274/Domain: annexin repeat homology <AX1>
F;275-346/Domain: annexin repeat homology <AX2>
F;358-430/Domain: annexin repeat homology <AX3>
F;434-505/Domain: annexin repeat homology <AX4>

Query Match          6.4%; Score 128; DB 2; Length 505;
Best Local Similarity 28.1%; Pred. No. 2.1;
Matches 55; Conservative 15; Mismatches 76; Indels 50; Gaps 9;

QY 139 ASYPPVPT--ATGPPNTSYMPGMPGSGISAYPSGYPPN--PS--GYPGCPY----- 182
Db 80 AGYPPVPGFGFGPPSAQQQVPYGYMYPFPGGNPSPMPSPYPYFGAPVPGQMPFGQQ 139
QY 183 PAGYPATNTSSQYSPQPPVTVV-----PSRDGTISDTI-----RASLISAV- 225
Db 140 PFGAYPGQPPVTVPGQPPVPLPGQQQVPVPSYPGSPGTVTPAVPPPTQFSGRGTTDAPG 199
QY 226 -----SDKLRWKEEMDGAQAEINALKRTEEDLKKGHKLEEMVTRLDQEVAEVDKNI 279
Db 200 FDPURDAEVLKRWKMGFGTQEQAIIDCLGRSN--KORQQILLISFKTAYGK----- 248
QY 280 ELLKKKDELSALEK 295
Db 249 DLIKOLKSELSGNFEK 264
```

```
RESULT 15
JC5368
dynactin 1 - mouse
N;Alternate names: p150 Glued
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1997 #sequence_revision 18-Jul-1997 #text_change 05-Nov-1999
C;Accession: JC5368
R;Jiang, W.; Weber, J.S.; Tokito, M.K.; Holzbaur, E.L.F.; Meisler, M.H.
Biochem. Biophys. Res. Commun. 231, 344-347, 1997
A;Title: Mouse p150Glued (dynactin 1) cDNA sequence and evaluation as a candidate for th
A;Reference number: JC5368; MUID:97223454; PMID:9070275
A;Accession: JC5368
A;Molecule type: mRNA
A;Residues: 1-1281 <JAN>
A;Cross-references: GB:U60312; NID:g2104494; PIDN:AAB57773.1; PID:g2104495
A;Experimental source: brain
C;Comment: This protein is a member of the oligomeric dynactin complex that is required
C;Genetics:
A;Gene: Dctn1
A;Map position: 6

Query Match          6.4%; Score 127.5; DB 2; Length 1281;
Best Local Similarity 24.5%; Pred. No. 7.3;
Matches 69; Conservative 46; Mismatches 104; Indels 63; Gaps 14;

QY 134 RPTVSASYPPTATGPPNTSYNMGPMGSGISAYSPGYPNPSGPGCPVPPAGYPATSS 193
DB 151 KPTRPAS---TGVAGPSSSLGPGSGSAGAGLSSEPSTPAQTPLA--APIIPTALTSP 204
QY 194 QYPSOPPVTYVGPSSRDGITS-----EDTIRASLISAVSDKLRW-----MK 234
DB 205 --GAAPPLPSKKEEGLRAQVRDLEELTURLKRSEDKAKLEKHKTGLEQVQWVK 262
QY 235 BEMDGAQAEL-----NALKRTBEDLKGKQKLEEM-----VTRLDQEVAE-----VD 276
DB 263 SKMQEQQADLQRLKEARKEAKELEARYNEEMADTADALEMATLDKEMAEERASLSQ 322
QY 277 KNIELLKKDBELSSALEKMQENQSENNDIDEVIIITAPLYKQILNLYAEENAIEDTIFVL 336
DB 323 QVEALKERVDLTLDLEILKAEIEBKGSQD---GAASSYQ--LKQLEFQNA-----RL 370
QY 337 GEALRGVIDLDVFLK--HV---RLLSRKQFOLRALMQKARK 373
DB 371 KDALVR--MDLSSSEKQBEVKVLQKLMEKKNQOELEVRQORER 411

RESULT 16
T05421
hypotheical protein F28A23.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C;Accession: T05421
R;Bevan, M.; Weichselgartner, M.; Fartmann, B.; Grandexath, K.; Dauner, D.; Herzl, A.; N
submitted to the Protein Sequence Database, October 1998
A;Reference number: Z15415
A;Accession: T05421
A;Molecule type: DNA
A;Residues: 1-279 <BEV>
A;Cross-references: ENBL:AL021961
A;Experimental source: cultivar Columbia; BAC clone F28A23
C;Genetics:
A;Map position: 4
A;Introns: 38/1; 73/1; 115/2; 139/2; 151/3
A;Note: F28A23.90

Query Match          6.4%; Score 127; DB 2; Length 279;
Best Local Similarity 36.0%; Pred. No. 1.2;
Matches 36; Conservative 7; Mismatches 27; Indels 30; Gaps 7;

QY 125 FGEEP-----PVFSRPTVSASY---PPYTATG-----PNTSYMP--GMPSGITS 163
DB 155 YGSAPSAFYAPHVPOYSAPPSASPTAPPYSGSLYPQVQYPPQSGYPYPASGYPPQPS 214
```

```
QY 164 AYP-----SGYPNPNSGYPGCPYPAGYPATTSQYPSQP 199
DB 215 AYPPTSTSGYPPIPSAYP--PPPTSSAYP--PPYPPQP 249

RESULT 17
G83385
hypotheical protein PA2075 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: G83385
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: G83385
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-553 <STO>
A;Cross-references: GB:AE004635; GB:AE004091; NID:g9948085; PIDN:AAG05463.1; GSPDB:GN0015
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA2075

Query Match          6.2%; Score 124; DB 2; Length 553;
Best Local Similarity 23.1%; Pred. No. 4.1;
Matches 52; Conservative 38; Mismatches 79; Indels 56; Gaps 9;

QY 103 YLHWKHPRSELLELIQIMVIFGEEPPVFSRPTVSAYPPYATGPPNTSYMPGMPGSI 162
DB 241 FVDNASHLREELADLAQML-----HPALHAA-----DATGSAN-----LPSPD 278
QY 163 SAYPSGYPNPSGYPGCPYPAGYPATTSQYPS-----QPPYTVGPSSRDGISE-D 215
DB 279 AAESSMPSPFS-----PLAAPAPEVSAVPSLSURLRVPRLAAGAGGANARAP 331
QY 216 TIRASLISAVSDKLR-----WRMKEEMDGAQAEL-----NALKRTBEDLKKGH 258
DB 332 TQAAAVIRQPEDEAHVDTDTALDIGNDTEHLLQDVQGTGLADLAGMAQQLSQQKQALQQ 391
QY 259 QKLEEMVTRL---DOEVAEVKNIELLKKDBELSSALEKMQENQS 300
DB 392 ESLESLOTQLHEKERQLOEKERQLRQWHLKQLQDDRQALERETEQS 436

RESULT 18
C86349
F8K7.4 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C;Accession: C86349
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86349
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1181 <STO>
A;Cross-references: GB:AE005172; NID:G5263313; PIDN:AAD41415.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match          6.2%; Score 123.5; DB 2; Length 1181;
```


C;Comment: Annexins undergo reversible, calcium-dependent binding to membrane phospholipid derstood.

C;Genetics:

A;Introns: 19/1; 58/2

A;Note: the list of introns is incomplete

C;Superfamily: annexin VII; annexin repeat homology

C;Keywords: alternative splicing; calcium binding; duplication; endonexin fold; glycoprotein

F:203-274/Domain: annexin repeat homology <AX1>

F:214-230/Region: endonexin fold #status predicted

F:275-346/Domain: annexin repeat homology <AX2>

F:286-302/Region: endonexin fold #status predicted

F:358-430/Domain: annexin repeat homology <AX3>

F:370-386/Region: endonexin fold #status predicted

F:434-505/Domain: annexin repeat homology <AX4>

F:445-461/Region: endonexin fold #status predicted

F:61,113/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.1%; Score 122.5; DB 1; Length 505;
Best Local Similarity 27.1%; Pred. No. 4.5;
Matches 55; Conservative 14; Mismatches 73; Indels 61; Gaps 10;

QY 125 FGEPPVFRSRTVSASYPPTATGPTNTSYMPGMPGSGISAYSPGYPNPSPGCPYP-- 182
Db 91 FGQPPPA-QQPV-----PSYGMYPGPGNPTSGMPS-----YPP-----YPGAPVPGQ 132

QY 183 -----PAGYPATTSSQYSPQPPVTVG-----PSRDGTISEDTI-----RA 219
Db 133 PMLPFGQQPPGVYGGPPMTYPCQSPVPPGQPPGYPSPGSGGTVTTPAVSPAQFGNKG 192

QY 220 SLISAV-----SDKLWRMKEEMDGAQAEALNALKRTEEDLKKGKLEEMVTRLDQEV 272
Db 193 TITDASGDFPLDAEVLVRKMGFGTDEQAIIDCLGRSN--KQKQIILLSPKTYAGK-- 248

QY 273 AEVDKNIELKKDEELSSALEK 295
Db 249 -----DLIKDKSELGNFEK 264

RESULT 22

C97702

cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)

C;Species: Rickettsia conorii

C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001

C;Accession: C97702

R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rickettsia conorii

A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A;Reference number: A97700; MUID:21442074; PMID:11557893

A;Accession: C97702

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1902 <KUR>

A;Cross-references: GB:AE006914; PIDN:AAL02557.1; PID:g15619052; GSPDB:GN00173

C;Genetics:

A;Gene: scal

Query Match 6.1%; Score 122.5; DB 2; Length 1902;
Best Local Similarity 23.9%; Pred. No. 24;
Matches 77; Conservative 46; Mismatches 110; Indels 89; Gaps 17;

QY 127 BEPPVF---SRPTVS-ASYPPTATGPTNTSY-----MGMPGSGISA--YPSGYPPN 172
Db 161 QPEIITITASSVTSPASNSFITAPNTPTITTSBEHYHTAGTPTSTPATPYQSTSDSK 220

QY 173 PSYGPVCPYPP-----AGP--YPATTSSQYSPQ--PPVTTVGSPSRDGTIS 213
Db 221 PNDSLGANTPPNINTNSKAVRELSPSSGPGQQAQVQSSQVSEVPKPTFPVPLIKSS 280

QY 214 EDTIRASLISAVS-----DKLWRMKEEMDGAQAEALNALKRTEEDLK 256
Db 281 TEIV-AGMVSNI SRVNMIGIKLAEVTQAIDTTDKKDKERLQKLTQTSTQTKTEKLG- 338

QY 257 GHQKLEENVTRLDQEVAEVDKNIELKKDEELSSALEKMNQSENNDIDEVIIP----- 311

C;Comment: Annexins undergo reversible, calcium-dependent binding to membrane phospholipid derstood.

C;Genetics:

A;Introns: 19/1; 58/2

A;Note: the list of introns is incomplete

C;Superfamily: annexin VII; annexin repeat homology

C;Keywords: alternative splicing; calcium binding; duplication; endonexin fold; glycoprotein

F:203-274/Domain: annexin repeat homology <AX1>

F:214-230/Region: endonexin fold #status predicted

F:275-346/Domain: annexin repeat homology <AX2>

F:286-302/Region: endonexin fold #status predicted

F:358-430/Domain: annexin repeat homology <AX3>

F:370-386/Region: endonexin fold #status predicted

F:434-505/Domain: annexin repeat homology <AX4>

F:445-461/Region: endonexin fold #status predicted

F:61,113/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.1%; Score 122; DB 2; Length 792;
Best Local Similarity 26.7%; Pred. No. 8.5;
Matches 50; Conservative 38; Mismatches 67; Indels 32; Gaps 9;

QY 215 DTIRASLISAVSDKLWRMKEEMDGAQAEALNALKR-----TEEDLKKGKLEEM- 264
Db 289 DVINLENLFSSQAALTHRLREEIDDKAQRALQKNNSSQTDNDMDMG--KKLKEMEKKV 347

QY 265 --VTRLDQEVAE---VDKNIELKKDEELSSALEKMNQSENNDIDEVIIPAPLYKQ 318
Db 348 NGVKDIDDEVEKSNIDKHLTRAMKLSFLSKRLKSTQEGDEBELKATNVPIDIGSL 407

QY 319 IINLYAEENAIEDTIFYLGEALRRGVIDL-----VFLKHVRLSRKQFQRLALMQKARK 373
Db 408 TDTKPEEN-IDDTV-----VSENALDIKSSEVVFAE--KDLSDENVQBEAETKT-K 457

QY 374 TAGLSDDL 380
Db 458 EASTSDL 464

RESULT 24

A70387

conserved hypothetical protein aq_1006 - Aquifex aeolicus

C;Species: Aquifex aeolicus

C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Feb-2001

C;Accession: A70387

R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Over

A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A;Reference number: A70300; MUID:98196666; PMID:9537320

A;Accession: A70387

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-978 <AQF>

A;Cross-references: GB:AB000718; NID:g2983504; PIDN:AAC07092.1; PID:g2983515; GB:AE00065;

A;Experimental source: strain VF5

C;Genetics:

A;Gene: aq_1006

C;Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Search completed: July 12, 2004, 08:30:39
Job time : 22.5223 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:32:51 ; Search time 299.393 Seconds
(without alignments)
396.939 Million cell updates/sec

Title: US-09-804-690-2
Perfect score: 1996
Sequence: 1 MMSKKYRDLTVRQTVNVIA.....FQIRALMQKARKTAGLSLDLY 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1279676 seqs, 311918243 residues

Total number of hits satisfying chosen parameters: 1279676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1996	100.0	381	9	US-09-804-690-2
2	1996	100.0	381	14	US-10-243-815A-3
3	1996	100.0	391	14	US-10-205-194-109
4	1996	100.0	391	14	US-10-376-564-1
5	1900.5	95.2	380	9	US-09-804-690-4
6	1900.5	95.2	390	14	US-10-243-815A-1
7	1900.5	95.2	390	14	US-10-053-975A-1
8	1900.5	95.2	390	14	US-10-376-564-2
9	1337.5	67.0	285	15	US-10-376-564-82
10	1337.5	67.0	307	15	US-10-264-049-2861
11	391	19.6	402	16	US-10-437-963-164064
12	390	19.5	87	14	US-10-029-386-28907
13	327	16.4	146	9	US-09-925-299-1257
14	327	16.4	146	10	US-09-925-299-1257
15	319	16.0	73	14	US-10-029-386-29129
					Sequence 2, Appli
					Sequence 3, Appli
					Sequence 109, Appl
					Sequence 1, Appli
					Sequence 4, Appli
					Sequence 1, Appli
					Sequence 2, Appli
					Sequence 82, Appl
					Sequence 2861, Ap
					Sequence 164064,
					Sequence 1257, Ap
					Sequence 1257, Ap
					Sequence 29129, A

16	11.1	221	12	US-10-424-599-239306	Sequence 239306,
17	190.5	296	9	US-09-801-368-386	Sequence 386, App
18	177.5	191	12	US-10-424-599-281387	Sequence 281387,
19	152	485	9	US-09-925-300-1664	Sequence 1664, Ap
20	150	466	16	US-10-408-765A-377	Sequence 377, App
21	146	7.5	12	US-10-115-123-453	Sequence 453, App
22	146	7.3	148	US-10-013-542-453	Sequence 453, App
23	146	7.3	176	US-10-106-698-5906	Sequence 5906, Ap
24	145	7.3	587	US-10-437-963-106668	Sequence 106668,
25	143	7.2	777	US-10-264-049-2971	Sequence 2971, Ap
26	142.5	7.1	983	US-10-112-944-435	Sequence 435, App
27	142.5	7.1	1021	US-10-408-765A-313	Sequence 313, App
28	138	6.9	258	US-10-424-599-176931	Sequence 176931,
29	138	6.9	278	US-10-425-114-68284	Sequence 68284, A
30	137.5	6.9	1251	US-10-112-944-881	Sequence 881, App
31	137	6.9	397	US-10-094-749-2983	Sequence 2983, App
32	135.5	6.8	827	US-10-149-310-68	Sequence 68, Appl
33	133.5	6.7	803	US-10-437-963-132841	Sequence 132841,
34	133	6.7	505	US-10-168-097A-56	Sequence 56, Appl
35	133	6.7	505	US-10-239-431A-36	Sequence 36, Appl
36	132.5	6.6	518	US-10-369-493-3240	Sequence 3240, Ap
37	131	6.6	358	US-10-437-963-153141	Sequence 153141,
38	131	6.6	1322	US-10-369-493-13040	Sequence 13040, A
39	131	6.6	2851	US-10-437-963-119072	Sequence 119072,
40	130.5	6.5	789	US-10-437-963-132698	Sequence 132698,
41	130.5	6.5	1121	US-10-437-963-196036	Sequence 196036,
42	130	6.5	787	US-10-104-047-3340	Sequence 3340, Ap
43	129.5	6.5	832	US-10-437-963-132792	Sequence 132792,
44	128.5	6.4	447	US-10-437-963-152072	Sequence 152072,
45	128.5	6.4	2137	US-10-437-963-118938	Sequence 118938,
46	128.5	6.4	2147	US-10-437-963-119058	Sequence 119058,
47	128	6.4	141	US-10-437-963-152071	Sequence 152071,
48	128	6.4	505	US-10-408-765A-539	Sequence 539, App
49	128	6.4	605	US-10-437-963-102875	Sequence 102875,
50	128	6.4	1025	US-10-437-963-118554	Sequence 118554,
51	127.5	6.4	2359	US-10-437-963-135813	Sequence 135813,
52	127.5	6.4	2359	US-10-437-963-119074	Sequence 119074,
53	127	6.4	860	US-10-437-963-132871	Sequence 132871,
54	126.5	6.3	416	US-10-043-487-282	Sequence 282, App
55	126.5	6.3	505	US-10-168-097A-12	Sequence 12, Appl
56	126.5	6.3	505	US-10-239-431A-32	Sequence 32, Appl
57	126.5	6.3	849	US-10-437-963-132731	Sequence 132731,
58	126	6.3	104	US-10-332-859-188	Sequence 188, App
59	125.5	6.3	979	US-10-408-765A-337	Sequence 337, App
60	125.5	6.3	1117	US-10-437-963-195619	Sequence 195619,
61	125.5	6.3	1945	US-10-437-963-195813	Sequence 195813,
62	125.5	6.3	2123	US-10-437-963-119074	Sequence 119074,
63	125.5	6.3	2579	US-10-437-963-119041	Sequence 119041,
64	124.5	6.2	371	US-09-284-320-2	Sequence 2, Appli
65	124.5	6.2	371	US-10-197-666A-4	Sequence 4, Appli
66	124.5	6.2	371	US-10-024-298A-41	Sequence 41, Appl
67	124.5	6.2	371	US-10-042-211A-41	Sequence 41, Appl
68	124.5	6.2	371	US-10-617-217A-41	Sequence 41, Appl
69	124.5	6.2	2057	US-10-437-963-119056	Sequence 119056,
70	124	6.2	336	US-10-437-963-201912	Sequence 201912,
71	124	6.2	667	US-10-424-539-162291	Sequence 162291,
72	124	6.2	2017	US-10-437-963-195768	Sequence 195768,
73	123.5	6.2	180	US-10-437-963-191496	Sequence 191496,
74	123.5	6.2	262	US-10-133-234A-5	Sequence 5, Appli
75	123.5	6.2	483	US-10-437-963-191498	Sequence 191498,
76	123.5	6.2	1794	US-10-437-963-118901	Sequence 118901,
77	123	6.2	744	US-10-282-122A-62864	Sequence 62864, A
78	123	6.2	832	US-10-437-963-132736	Sequence 132736,
79	123	6.2	987	US-10-437-963-119086	Sequence 119086,
80	122.5	6.1	367	US-10-437-963-146546	Sequence 146546,
81	122.5	6.1	501	US-10-168-097A-46	Sequence 46, Appl
82	122.5	6.1	501	US-10-239-431A-35	Sequence 35, Appl
83	122.5	6.1	753	US-10-369-493-2333	Sequence 2333, Ap
84	122.5	6.1	816	US-10-437-963-132787	Sequence 132787,
85	122.5	6.1	831	US-10-437-963-132867	Sequence 132867,
86	122.5	6.1	1017	US-10-437-963-159865	Sequence 159865,
87	122.5	6.1	1253	US-10-334-143-2	Sequence 2, Appli
88	122.5	6.1	2034	US-10-437-963-195746	Sequence 195746,

89 122 6.1 813 16 US-10-437-963-132798 Sequence 132798,
90 122 6.1 889 16 US-10-437-963-197045 Sequence 197045,
91 121.5 6.1 727 16 US-10-437-963-132675 Sequence 132675,
92 121.5 6.1 827 16 US-10-437-963-132839 Sequence 132839,
93 121.5 6.1 832 16 US-10-437-963-132728 Sequence 132728,
94 121.5 6.1 1023 16 US-10-437-963-119007 Sequence 119007,
95 121.5 6.1 1780 16 US-10-437-963-171376 Sequence 171376,
96 121 6.1 204 12 US-10-424-599-164990 Sequence 164990,
97 121 6.1 306 16 US-10-437-963-146544 Sequence 146544,
98 121 6.1 808 16 US-10-437-963-132681 Sequence 132681,
99 121 6.1 811 16 US-10-437-963-117066 Sequence 117066,
100 121 6.1 812 16 US-10-437-963-185099 Sequence 185099,

ALIGNMENTS

RESULT 1
US-09-804-690-2
; Sequence 2, Application US/09804690
; Patent No. US20020034743A1
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/804,690
; FILING DATE: 12-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/146,187
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-804-690-2

Query Match 100.0%; Score 1996; DB 9; Length 381;
Best Local Similarity 100.0%; Pred. No. 3.1e-141;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MMSKYKRDLTVRQTVNVIAWKLPVLDYVNDGSSRELNLVTGTPVYRGNTYNI 60
Db 1 MMSKYKRDLTVRQTVNVIAWKLPVLDYVNDGSSRELNLVTGTPVYRGNTYNI 60
Qy 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDWKHPRSELELIQI 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDWKHPRSELELIQI 120

Qy 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSIAYPSGYPPNPSGYGCP 180
Db 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSIAYPSGYPPNPSGYGCP 180
Qy 181 YPPAGPYPATSSQYPSQPPVTVGFSRDGTISEDTRASLISAVSKLRWRKEEMDGA 240
Db 181 YPPAGPYPATSSQYPSQPPVTVGFSRDGTISEDTRASLISAVSKLRWRKEEMDGA 240
Qy 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDELSALEKMNQS 300
Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDELSALEKMNQS 300
Qy 301 ENNDIDEVITPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 301 ENNDIDEVITPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Qy 361 QFQLRALMQKARKTAGLSDLV 381
Db 361 QFQLRALMQKARKTAGLSDLV 381

RESULT 2

US-10-243-815A-3
; Sequence 3, Application US/10243815A
; Publication No. US20030099988A1
; GENERAL INFORMATION:
; APPLICANT: La Brie, Samuel T.; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR SUPPRESSOR
; FILE REFERENCE: PF-0199-2 DIV
; CURRENT APPLICATION NUMBER: US/10/243,815A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 09/216,387
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: US 08/786,999
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank ID No. US20030099988A1 g1330330
US-10-243-815A-3

Query Match 100.0%; Score 1996; DB 14; Length 381;
Best Local Similarity 100.0%; Pred. No. 3.1e-141;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMSKYKRDLTVRQTVNVIAWKLPVLDYVNDGSSRELNLVTGTPVYRGNTYNI 60
Db 1 MMSKYKRDLTVRQTVNVIAWKLPVLDYVNDGSSRELNLVTGTPVYRGNTYNI 60
Qy 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDWKHPRSELELIQI 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDWKHPRSELELIQI 120
Qy 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSIAYPSGYPPNPSGYGCP 180
Db 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSIAYPSGYPPNPSGYGCP 180
Qy 181 YPPAGPYPATSSQYPSQPPVTVGFSRDGTISEDTRASLISAVSKLRWRKEEMDGA 240
Db 181 YPPAGPYPATSSQYPSQPPVTVGFSRDGTISEDTRASLISAVSKLRWRKEEMDGA 240
Qy 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDELSALEKMNQS 300
Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDELSALEKMNQS 300
Qy 301 ENNDIDEVITPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 301 ENNDIDEVITPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360

Db 301 ENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFVLGEALRRGVLDLDVFLKHVRLLSRK 361

Qy 361 QFQLRALMQARKTAGLSLDLY 381
 |||||

Db 361 QFQLRALMQARKTAGLSLDLY 381

RESULT 3

US-10-205-194-109

; Sequence 109, Application US/10205194

; Publication No. US20030134301A1

; GENERAL INFORMATION:

; APPLICANT: Warner-Lambert Company

; APPLICANT: Lee, Kevin

; APPLICANT: Dixon, Alistair

; APPLICANT: Brooksbank, Robert

; APPLICANT: Pinnoch, Robert

; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain

; FILE REFERENCE: WL-A-018201

; CURRENT APPLICATION NUMBER: US/10/205,194

; CURRENT FILING DATE: 5200-07-24

; PRIOR APPLICATION NUMBER: GB 0118354.0

; PRIOR FILING DATE: 2001-07-27

; NUMBER OF SEQ ID NOS: 177

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 109

; LENGTH: 391

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

; OTHER INFORMATION: Tumor susceptibility protein TSG101

US-10-205-194-109

Query Match 100.0%; Score 1996; DB 14; Length 391;

Best Local Similarity 100.0%; Pred. No. 3.2e-141;

Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 MMSKYYKRDLTVRGVNVNVIAMYKDLKPVLDSYVFNDSGSSRELNVLTGTIPVYRGNINYI 60

Db 11 MMSKYYKRDLTVRGVNVNVIAMYKDLKPVLDSYVFNDSGSSRELNVLTGTIPVYRGNINYI 70

Qy 61 PICLWLDDTYPNPPICFVKPTSSMTIKGKHVDANGKIYIPLYLHDWKHPRSELLELIQI 120

Db 71 PICLWLDDTYPNPPICFVKPTSSMTIKGKHVDANGKIYIPLYLHDWKHPRSELLELIQI 130

Qy 121 MIVIFGEPPVFSRPTVSASYPPTYATGPPNTSYMPGMSGISAYPSGYVPPNPSGYGCP 180

Db 131 MIVIFGEPPVFSRPTVSASYPPTYATGPPNTSYMPGMSGISAYPSGYVPPNPSGYGCP 190

Qy 181 YPPAGPYPATTSQYPSQPPVTTVGPSPRDGTISEDTRASLISAVSDKLRWRMKEEMDGA 240

Db 191 YPPAGPYPATTSQYPSQPPVTTVGPSPRDGTISEDTRASLISAVSDKLRWRMKEEMDGA 250

Qy 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQAEVAVDKNIELLKKDEELSSALERWQNS 300

Db 251 QAEINALKRTEEDLKKGHQKLEEMVTRLDQAEVAVDKNIELLKKDEELSSALERWQNS 310

Qy 301 ENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFVLGEALRRGVLDLDVFLKHVRLLSRK 360

Db 311 ENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFVLGEALRRGVLDLDVFLKHVRLLSRK 370

Qy 361 QFQLRALMQARKTAGLSLDLY 381

Db 371 QFQLRALMQARKTAGLSLDLY 391

RESULT 4

US-10-376-564-1

; Sequence 1, Application US/10376564

; Publication No. US20030180302A1

; GENERAL INFORMATION:

; APPLICANT: wolf, Eckhard

; APPLICANT: Werner, Sabine

; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/804,690
; FILING DATE: 12-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/146,187
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-804-690-4

Query Match 95.2%; Score 1900.5; DB 9; Length 380;
Best Local Similarity 94.5%; Pred. No. 4.4e-134;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MMSKYKRDLTVRQTVNVVIAMVKDLKPVLDYVFNDSGSSRELNVLTGTIPVYRGNTYNI 60
Db 1 MVSXYKYRDLTVRETQVNVITLYKDLKPVLDYVFNDSGSSRELNVLTGTIPVYRGNTYNI 60

QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDMKHPRSELLELIQI 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDMKHPRSELLELIQV 120

QY 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180
Db 121 MIVVFGDEPPVFSRPP-ISASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 179

QY 181 YPPAGPYPATSSQYPSQPPVTVTPGSRDGTISDTIRASLISAVSDKLWRMKEEMDGA 240
Db 180 YPPGGPYPATSSQYPSQPPVTVTPGSRDGTISDTIRASLISAVSDKLWRMKEEMDRA 239

QY 241 QAEINALKRTEEDLKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNOS 300
Db 240 QAEINALKRTEEDLKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNOS 299

QY 301 ENNDIDEVITPTAPLYKQILNLVAENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 300 ENNDIDEVITPTAPLYKQILNLVAENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 359

QY 361 QFQLRALMOKARKTAGLSLDLY 381
Db 360 QFQLRALMOKARKTAGLSLDLY 380

RESULT 6
US-10-243-815A-1
; Sequence 1, Application US/10243815A
; Publication No. US2003009988A1
; GENERAL INFORMATION:
; APPLICANT: La Brie, Samuel T.; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR SUPPRESSOR
; FILE REFERENCE: PF-0199-2 DIV
; CURRENT APPLICATION NUMBER: US/10/243,815A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 09/216,387

; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: US 08/786,999
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US2003009988A1 609476CD1
US-10-243-815A-1

Query Match 95.2%; Score 1900.5; DB 14; Length 390;
Best Local Similarity 94.5%; Pred. No. 4.5e-134;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MMSKYKRDLTVRQTVNVVIAMVKDLKPVLDYVFNDSGSSRELNVLTGTIPVYRGNTYNI 60
Db 11 MVSXYKYRDLTVRETQVNVITLYKDLKPVLDYVFNDSGSSRELNVLTGTIPVYRGNTYNI 70

QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDMKHPRSELLELIQI 120
Db 71 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDMKHPRSELLELIQV 130

QY 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180
Db 131 MIVVFGDEPPVFSRPP-ISASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 189

QY 181 YPPAGPYPATSSQYPSQPPVTVTPGSRDGTISDTIRASLISAVSDKLWRMKEEMDGA 240
Db 190 YPPGGPYPATSSQYPSQPPVTVTPGSRDGTISDTIRASLISAVSDKLWRMKEEMDRA 249

QY 241 QAEINALKRTEEDLKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNOS 300
Db 250 QAEINALKRTEEDLKGHQKLEEMVTRLDQEAEDVKNIELKKKDELSALEKMNOS 309

QY 301 ENNDIDEVITPTAPLYKQILNLVAENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 310 ENNDIDEVITPTAPLYKQILNLVAENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 369

QY 361 QFQLRALMOKARKTAGLSLDLY 381
Db 370 QFQLRALMOKARKTAGLSLDLY 390

RESULT 7
US-10-053-975A-1
; Sequence 1, Application US/10053975A
; Publication No. US20030138839A1
; GENERAL INFORMATION:
; APPLICANT: LI, LIMIN
; APPLICANT: COHEN, STANLEY N.
; TITLE OF INVENTION: "Mammalian Tumor Susceptibility Gene
; TITLE OF INVENTION: Products and Their Uses"
; FILE REFERENCE: STAN-216
; CURRENT APPLICATION NUMBER: US/10/053,975A
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,763
; PRIOR FILING DATE: 2001-01-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-053-975A-1

Query Match 95.2%; Score 1900.5; DB 14; Length 390;
Best Local Similarity 94.5%; Pred. No. 4.5e-134;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;


```
Db 254 FLKHVLLSRKQFOLRALMOKARKTAGLSDIY 285

RESULT 10
US-10-264-049-2861
; Sequence 2861, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2861
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (24)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2861

Query Match 67.0%; Score 1337.5; DB 15; Length 307;
Best Local Similarity 95.2%; Pred. No. 5.1e-92;
Matches 259; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Qy 110 PRSLLLELIQIMIVIFGEEPPVRSRPTVSASYPYATGTPNTSYMPGMSGISAYPSGY 169
Db 37 PQSDLLGLIQIMIVFGBEPVFRP-ISASYPYQATGTPNTSYMPGMSGISAYPSGY 95

Qy 170 PNPSPGPGCPYPAGYPATTSQYPSQPPVTTVTGFPSRDGTISEDITRASLISAVSDKL 229
Db 96 PNPSPGPGCPYPGPGYPATTSQYPSQPPVTTVTGFPSRDGTISEDITRASLISAVSDKL 155

Qy 230 RWRKKEMDGAQELNALKETEEDLKKGHQKLEEMVTRLDQEAEDVNDKNIELKKDBEL 289
Db 156 RWRKKEMDGAQELNALKETEEDLKKGHQKLEEMVTRLDQEAEDVNDKNIELKKDBEL 215

Qy 290 SSALKEMENSENDIDEVIPTAPLYKQIINLYAENNAEDTTFYLGEALRRGVIDLVDV 349
Db 216 SSALKEMENSENDIDEVIPTAPLYKQIINLYAENNAEDTTFYLGEALRRGVIDLVDV 275

Qy 350 FLKHVLLSRKQFOLRALMOKARKTAGLSDIY 381
Db 276 FLKHVLLSRKQFOLRALMOKARKTAGLSDIY 307

RESULT 11
US-10-437-963-164064
; Sequence 164064, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 164064
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_62C.1.pap
US-10-437-963-164064

Query Match 19.6%; Score 391; DB 16; Length 402;
Best Local Similarity 30.1%; Pred. No. 6.3e-21;
Matches 108; Conservative 65; Mismatches 148; Indels 38; Gaps 11;

Qy 22 YKDLKPYLDSYFNDGSSRELNLGTGTPVRYGNIYNIPICLWLLDTPYNNPPICFVKP 81
Db 56 FPSLHPKAALFTHNDGRAHLQADGTPIHHAGASYNLPAVLWLPPEYPRSPPLVFLSP 115

Qy 82 TSSMTIKTKGH--VDANGKI-YLPYLHDWKHPRSELELELIQIMIVIFGEEPPVRSRPTVS 138
Db 116 TRDMVIKP-HHPLVDRSGLVANAPYLRSWVFPSSNLVDLVRSLSHLFLGLDPLPFR---- 170

Qy 139 ASYPYPYTATGTPNTSYMPGMSGISAYPSGYPNPSPGPGCPYPAGYPATTSQYPSQ 198
Db 171 SPNPFPSPSPPP----IPATP-----LPRVHPSSSSS-----PSPSPYFPASPQLAAR 215

Qy 199 PPVTTVGPSRD-GTISEDITRASLIS---AVSKLRWKEEMDGAQELNALKRTBEDL 254
Db 216 PP-----PTEDPAEVYKRNAIAKLVDMAADAATLRPVREAEDVTLFAMQATLRSRGEV 270

Qy 255 KGHQKLEEMVTRLDQEAEDVNDKNIELKKDBELSSALEKMEQSENNDIDEVIITAP 314
Db 271 SDGVRKMGEEKALERLQVMMATDLME-----AWMENTKGAAGDTEADEAETADV 324

Qy 315 LYQIINLYAENNAEDTTFYLGEALRRGVIDLVDLFLKHVLLSRKQFOLRALMOKARK 373
Db 325 LSKQMLECTAADLAEDTTFYALDKAIQEGSVPPGYLRSVRALAREQFFQFVLSTKYNK 383

RESULT 12
US-10-029-386-28907
; Sequence 28907, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28907
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR11.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.83
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.6
; OTHER INFORMATION: SWISSPROT HIT: Q99816, EVALUAE 3.00e-38
US-10-029-386-28907

Query Match 19.5%; Score 390; DB 14; Length 87;
Best Local Similarity 100.0%; Pred. No. 9.5e-22;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
,
,
, TYPE: PRT
, ORGANISM: Homo sapiens
, FEATURE:
, OTHER INFORMATION: MAP TO CHR11.1
, OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
, OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
,
```

```
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN HEPATOCYTE, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: Q99816, EVALUE 4.00e-31
US-10-029-386-29129

Query Match          16.0%; Score 319; DB 14; Length 73;
Best Local Similarity 81.8%; Pred. No. 1.6e-16;
Matches 54; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 44 NLGTGTPVRYRGNINIPICLWLDTPYNNPPICFVKPTSSMTIKTKGKVDANGKIYLPY 103
Db 5 NRVTEISPSGNTYINIPICLWLDTPYNNPPICFVKPTSSMTIKTKGKVDANGKIYLPY 64

QY 104 LHDWKH 109
Db 65 LHEWKH 70

RESULT 16
US-10-424-599-239306
; Sequence 239306, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 239306
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58119C.1.pep
US-10-424-599-239306

Query Match          11.1%; Score 221; DB 12; Length 231;
Best Local Similarity 27.9%; Pred. No. 1.6e-08;
Matches 61; Conservative 30; Mismatches 78; Indels 50; Gaps 8;

QY 12 VRQ-TVNVAMYKDLKPVLDSYVNDGSSRELNLGTGTPVRYRGNINIPICLWLDTPY 70
Db 34 IRQHLVALTAFPSLEPKTASFTHNDGSRVNLQADGTIEMTFQGVTYNIPVVIWLMESY 93

QY 71 PYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLDHDKHPRSELLELIQIMIVIFGEEP 129
Db 94 PRHPCCVYNPTDMLIKRPHPVNFSGLVSVYLPQWYTP-----GDDS 138

QY 130 PVFSRPTVSASYPYATGP-----PNTSY-----MPCMPSGISAYPSGYPN- 172
Db 139 VPQLRP-----FPARPLRGLLRGHRVHPLAPRRSHHCLPFPAPRTPSFTHNDGSRVNL 193

QY 173 -----PSGYPGCPFPAGPYATTSSQYPSQPPVTV 204
Db 194 LQADGTIEMTFQGVTY-----NVPVVISPSPCFAPSSRV 228

RESULT 17
US-09-801-368-386
; Sequence 386, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin

; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 386
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-09-801-368-386

Query Match          9.5%; Score 190.5; DB 9; Length 296;
Best Local Similarity 23.6%; Pred. No. 4.4e-06;
Matches 71; Conservative 54; Mismatches 125; Indels 51; Gaps 13;

QY 7 YRD-LTVRQTVNVAMYKDLKPVLDSYVNDGSSRELNLGTGTPVRYRGNI-YNIPIC 63
Db 25 YNDGRTTFHDSLALDNLHPSLRPRTRVFTHSDDGTQLLLSIYGTISTGDSGSSPHSIPVI 84

QY 64 LWLDDTPYNNPPICT-----FVKPTSSMTIKTKGKVDANGKIYLPYLDHDKHPRSELLEL 117
Db 85 MWVPSMYVKKPPFISINLENFDMNTISSSLPIQEIYDTSNGWIALPILHCDPAAMNLMV 144

QY 118 IQIMIVIFGEEPPVFSRPTVSASYPYATP-----GPNTSYMPCMPSGISAYPSGYPNPS 174
Db 145 VQELMSLL-HEPPDQAP-----SLPKPNTLOQRQNTPELPKPKS-----PHLKPLP- 194

QY 175 GYPCPYPPAGPYATTSSQYPSQPPVTVVPSRD-----GTISDITIRASLISAVSD 227
Db 195 -----PPPPPPQASNALDMLMDMN-TDISPTNHHMLQNLQTVVNNELYRED-VDYVAD 245

QY 228 KLRWR---WKEENDGQAELNALKRTEEDLKKGHQKLEEMVTELDQEAEDVKNIELKK 284
Db 246 KILTRQTVMQE-----SIARFHEIIADKNHLRAVEQAIEQTMHSLNAQIDVLR 295

QY 285 K 285
Db 296 K 296

RESULT 18
US-10-424-599-283387
; Sequence 283387, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 283387
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
```


; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 106668
; LENGTH: 587
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_11091C.1.pep
US-10-437-963-106668

Query Match 7.3%; Score 145; DB 16; Length 587;
Best Local Similarity 26.6%; Pred. No. 0.029;
Matches 51; Conservative 19; Mismatches 64; Indels 58; Gaps 8;

QY	71	PYNPPICFVKPTSSMTIKTKHVDANGKIYLPYLHDKHPRSELLELIQIMIVIFGEPP	130
DB	236	PISSCCAPPTSTTTTGAPYPPPPAMASYPPL-----PS	271
QY	131	VFSRPTVS-----ASYPP-----YTATGPNNTSYMPGMPG--ISAYP-----	166
DB	272	LSATPSASLYPPPPSSYPPPPPPPPHVTQSPAPNSSYPPPPPSQYIAGYPPPPPS	331
QY	167	SGYPNPGYGCYP-----PAGYPATTSSQSPQPPVTTVGPSRDGTISEDITRAS	220
DB	332	NFYPPPPAGYAPSPPTSTYPPPPPPESASSQYP--PPLPSAPCCDRSV--DRALPS	387
QY	221	LISAVSDKLRWR	232
DB	388	YMSPPREADGWR	399

RESULT 25
US-10-264-049-2971
; Sequence 2971, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2971
; LENGTH: 777
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (186)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2971

Query Match 7.2%; Score 143; DB 15; Length 777;
Best Local Similarity 25.9%; Pred. No. 0.059;
Matches 50; Conservative 32; Mismatches 85; Indels 26; Gaps 7;

QY	122	IVIFGEE-PPVFSRPTVSASYPPTATGPNNTSYMPGMPG-----GISAYPSGYPNPSG	175
DB	435	VVIVPEGULPPLPPPPSGATPPPIAPTGPPTAS--PPVPAKEEPPELPAPGFLPPP--	490
QY	176	YGPCYPYPAGYPATTSSQSPQPPVTTVGPSRDGTISEDITRASLISAVSDKLRWKE	235
DB	491	-----PPPPPVPGVTLPPQLVPEGTPGGGGPAALEDLAVININSDEE-----EE	538
QY	236	EMDGAQAEINALKRTEEDLKGHQKLEEMVRLDQEVAVDKNIELLKKKDEELSALEK	295

Db	539	EEEEEEEEEEEEEDDFREEDDEDEDEYFEEEEEEEFEEFE-----EEGEELEE-EE	594
QY	296	MENQSENNDIDEV	308
DB	595	EEDEEEEEEELEEV	607

Search completed: July 12, 2004, 08:52:18
Job time : 300.393 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 38.5506 Seconds

(without alignments)
3118.303 Million cell updates/sec

Title: US-09-804-690-2

Perfect score: 1996

Sequence: 1 MMSKYYKDLVTRQIVNVIA.....FOLRALMOKARKTAGLSLDLY 381

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mbc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1956	98.0	391	11	Q7tse5 rattus norv
2	1894.5	94.9	390	4	Q9bum5 homo sapien
3	1813.5	90.9	392	13	Q918g8 chelonias my
4	964	48.3	408	5	Q9vva7 drosophila
5	725	36.3	425	5	Q76258 caenorhabdi
6	541.5	25.1	249	5	Q8mqz0 drosophila
7	505	25.3	114	13	Q7t2m2 ameiusus ne
8	440	22.0	398	10	Q91hgh8 arabidopsis
9	415	20.8	379	4	Q81x04 homo sapien
10	412	20.6	174	11	Q08761 mus musculu
11	408	20.4	249	11	Q7TQD3 mus musculu
12	389.5	19.5	368	10	Q9FFV6 arabidopsis
13	361	18.1	177	4	Q9ffiy6 arabidopsis
14	361	18.1	341	4	Q9NUX7 homo sapien
15	297	14.9	378	3	Q873m7 yarrowia li
16	293.5	14.7	385	3	P87279 saccharomyc

17	283	14.2	111	11	Q8BU96
18	190.5	9.5	296	3	Q8NIM6
19	167.5	8.4	376	3	P78998
20	151	7.6	148	4	Q9NTQ8
21	147	7.4	503	11	Q921F1
22	146	7.3	148	4	Q9NZ81
23	145	7.3	1130	4	Q81ZL8
24	143	7.2	746	4	Q9BU60
25	143	7.2	1052	4	Q96FT1
26	142.5	7.1	1021	4	O15451
27	142.5	7.1	1284	4	O15450
28	142	7.1	995	5	Q9NUJ9
29	141.5	7.1	485	3	Q9UTP6
30	141	7.1	485	13	Q804G3
31	140.5	7.0	3326	12	O7T591
32	138	6.9	809	13	O7ZVN7
33	137.5	6.9	247	10	Q945X9
34	137	6.9	397	4	Q8NEZ2
35	137	6.9	397	4	Q96DL9
36	135.5	6.8	463	11	Q922A2
37	135.5	6.8	827	3	O60130
38	135	6.8	137	11	Q9CQJ5
39	134.5	6.7	840	10	Q84SJ9
40	134	6.7	1032	10	Q7XXN3
41	133.5	6.7	415	11	Q8CCV9
42	133	6.7	201	16	O86316
43	132.5	6.6	319	11	Q8BP27
44	132.5	6.6	472	3	O59907
45	132	6.6	198	16	Q7U0B9
46	132	6.6	3288	12	Q7T5D9
47	131.5	6.6	303	11	Q9D116
48	131.5	6.6	319	11	Q9D0D7
49	131.5	6.6	319	11	Q9D4W4
50	131.5	6.6	463	11	Q8BP75
51	131	6.6	358	10	Q942Z3
52	130.5	6.5	463	11	Q8VIN2
53	130.5	6.5	1099	10	Q8W362
54	130.5	6.5	1099	10	Q7XDA1
55	130	6.5	609	17	Q8TXA4
56	130	6.5	787	4	Q8NAB1
57	130	6.5	1006	11	O62901
58	130	6.5	1253	4	Q86SQ0
59	129	6.5	695	5	Q86NU9
60	129	6.5	838	3	Q871B7
61	129	6.5	838	5	Q86NK0
62	129	6.5	866	5	Q8T8N7
63	129	6.5	924	5	Q9GYS7
64	129	6.5	925	5	Q9BIC2
65	128.5	6.4	303	11	Q8R3W0
66	128.5	6.4	1053	10	Q8LLY7
67	128.5	6.4	1053	10	Q7XVL5
68	127.5	6.4	483	13	Q7T391
69	127.5	6.4	505	4	O7Z746
70	127	6.4	279	10	O49490
71	127	6.4	471	10	Q9LDB31
72	127	6.4	581	10	Q7XVH4
73	127	6.4	604	5	O7X339
74	126.5	6.3	331	10	Q93V84
75	126.5	6.3	345	11	Q9ESF4
76	126.5	6.3	349	10	Q9LUC0
77	126	6.3	104	11	Q8K353
78	126	6.3	754	13	O8J777
79	125.5	6.3	280	9	Q856F3
80	125.5	6.3	332	13	Q8QCD9
81	125.5	6.3	461	12	Q9J861
82	125.5	6.3	979	5	Q9W5A4
83	125.5	6.3	997	10	Q8R225
84	125.5	6.3	1053	10	Q94562
85	125.5	6.3	1053	10	Q8H4H9
86	125.5	6.3	1114	5	Q8MT16
87	125.5	6.3	1114	5	Q7VZM4
88	125.5	6.3	1153	5	Q8IRY0
89	125.5	6.3	2123	10	Q7XES9

Q8bu96	mus musculu
Q8nim6	saccharomyc
P78998	saccharomyc
Q9ntq8	homo sapien
Q921f1	mus musculu
Q9nz81	homo sapien
Q81z18	homo sapien
Q9bu60	homo sapien
Q96ft1	homo sapien
O15451	homo sapien
O15450	homo sapien
Q9nj19	aequipecten
Q9utp6	schizosacch
Q804g3	brachydanio
O7t591	simian herp
O7zvn7	brachydanio
Q945x9	arabidopsis
Q8nez2	homo sapien
Q96dl9	homo sapien
Q922a2	mus musculu
O60130	schizosacch
Q9cqj5	mus musculu
Q84sj9	oryza sativ
Q7xxn3	oryza sativ
Q8ccv9	mus musculu
O86316	mycobacteri
Q8bp27	mus musculu
O59907	neurospora
Q7u0b9	mycobacteri
Q7t5d9	simian herp
Q9d116	mus musculu
Q9d0d7	mus musculu
Q9d4w4	mus musculu
Q8bp75	mus musculu
Q942z3	oryza sativ
Q8vin2	rattus norv
Q8w362	oryza sativ
Q7xdal	oryza sativ
Q8txa4	methanopyru
Q8nab1	homo sapien
Q62901	rattus norv
Q86sq0	homo sapien
Q86nj9	caenorhabdi
Q871b7	neurospora
Q86nk0	caenorhabdi
Q8t8n7	caenorhabdi
Q9gys7	caenorhabdi
Q9bic2	caenorhabdi
Q8r3w0	mus musculu
Q8lly7	oryza sativ
Q7xvl5	oryza sativ
Q7t391	brachydanio
Q7z746	homo sapien
Q49490	arabidopsis
Q9ld31	crypthecodi
Q7xvh4	oryza sativ
Q97339	phascocolin
Q93v84	arabidopsis
Q9esf4	mus musculu
Q9luc0	arabidopsis
Q8k353	mus musculu
Q8j777	xenopus lae
Q856f3	mycobacteri
Q8qcd9	gallus gall
Q9j861	spodoptera
Q9w5a4	drosophila
Q8r225	oryza sativ
Q94562	oryza sativ
Q8h4h9	oryza sativ
Q8mt16	drosophila
Q7vzm4	drosophila
Q8iry0	drosophila
Q7xes9	oryza sativ

```
90 125 6.3 508 13 093447 093447 oryzias lat
91 125 6.3 1053 10 07XKP7 07XKP7 oryza sativ
92 124.5 6.2 371 4 07Z429 07Z429 homo sapien
93 124.5 6.2 397 11 08CHS8 08chs8 mus musculu
94 124.5 6.2 739 10 08L7F7 08l7f7 arabidopsis
95 124.5 6.2 846 13 07ZTK5 07ztk5 xenopus lae
96 124.5 6.2 847 11 08R3S8 08r3s8 mus musculu
97 124.5 6.2 966 3 001385 001385 neurospora
98 124.5 6.2 977 3 08X005 08x005 neurospora
99 124.5 6.2 1269 10 07XR32 07xr32 oryza sativ
100 124 6.2 256 2 09L595 09l595 streptococc
```

ALIGNMENTS

```
RESULT 1
Q7TS55 PRELIMINARY; PRT; 391 AA.
AC Q7TS55; 2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Tumor susceptibility gene 101 protein.
GN TSG101.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP STRAIN=F344;
RC Burdorf S., Scheidtmann K.H.;
RT "Functional interaction between AATF, TSG101 and AR.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY293306; AAP45008.1; -.
SQ SEQUENCE 391 AA; 44096 MW; F6274A94DBD46CA3 CRC64;
```

```
Query Match 98.0%; Score 1956; DB 11; Length 391;
Best Local Similarity 97.6%; Pred. No. 4.1e-122;
Matches 372; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 MMSKYKRDLTVRQTVNVIAMYKDLKPVLDYVFNDSGSSRELNLGTIPVRYGNLYNI 60
DB 11 MMSKYKRDLTVRQTVNVIAMYKDLKPVLDYVFNDSGSSRELNLGTIPVRYGNLYNI 70
QY 61 PICLWLDITYPNPPICFVKPTSSMTIKTGHVDANGKIYLPYLDHMKHPRSELLELIQI 120
DB 71 PICLWLDITYPNPPICFVKPTSSMTIKTGHVDANGKIYLPYLDHMKHPRSELLELIQI 130
QY 121 MIVIFGEEPPVFSRPTVSAPYPTATGPPNTSYMPGPGISAYPSGYPNPSGYPGCP 180
DB 131 MIVIFGEEPPVFSRPTVSAPYPTATGPPNTSYMPGPGISAYPSGYPNPSGYPGCP 190
QY 181 YPPAGPYPATSSQYPSQPPVTTVGSRDGTISEDTIRASLISAVSKLRMRKEEMDGA 240
DB 191 YPPAGPYPATSSQYPSQPPVTTVGSRDGTISEDTIRASLISAVSKLRMRKEEMDGA 250
QY 241 QAEINALKRTEEDLKKGHQLKEEMVTRLDQVAEVDKNIELKKKDEELSSALEKMNQS 300
DB 251 QAEINALKRTEEDLKKGHQLKEEMVTRLDQVAEVDKNIELKKKDEELSSALEKMNQS 310
QY 301 ENNDIDEVIITAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
DB 311 ENNDIDEVIITAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 370
QY 361 QFQRLALMQARKTAGLSLDLY 381
DB 371 QFQRLALMQARKTAGLSLDLY 391
```

```
RESULT 2
Q9BUM5
```

```
ID Q9BUM5 PRELIMINARY; PRT; 390 AA.
AC Q9BUM5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Tumor susceptibility gene 101.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Eve;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR GO; GO:0004840; Fubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; P-ubiquitin cycle; IEA.
DR InterPro; IPR000883; Tsg101.
DR InterPro; IPR000608; UBQ conjugat.
DR Pfam; PF05743; Tsg101; 1.
DR SMART; SM00212; UBCC; 1.
SQ SEQUENCE 390 AA; 43910 MW; 96BB2A4FC22DF16A CRC64;
```

```
Query Match 94.9%; Score 1894.5; DB 4; Length 390;
Best Local Similarity 94.2%; Pred. No. 5.1e-118;
Matches 359; Conservative 11; Mismatches 10; Indels 1; Gaps 1;
```

```
QY 1 MMSKYKRDLTVRQTVNVIAMYKDLKPVLDYVFNDSGSSRELNLGTIPVRYGNLYNI 60
DB 11 MMSKYKRDLTVRQTVNVIAMYKDLKPVLDYVFNDSGSSRELNLGTIPVRYGNLYNI 70
QY 61 PICLWLDITYPNPPICFVKPTSSMTIKTGHVDANGKIYLPYLDHMKHPRSELLELIQI 120
DB 71 PICLWLDITYPNPPICFVKPTSSMTIKTGHVDANGKIYLPYLDHMKHPRSELLELIQI 130
QY 121 MIVIFGEEPPVFSRPTVSAPYPTATGPPNTSYMPGPGISAYPSGYPNPSGYPGCP 180
DB 131 MIVIFGEEPPVFSRPTVSAPYPTATGPPNTSYMPGPGISAYPSGYPNPSGYPGCP 189
QY 181 YPPAGPYPATSSQYPSQPPVTTVGSRDGTISEDTIRASLISAVSKLRMRKEEMDGA 240
DB 190 YPPAGPYPATSSQYPSQPPVTTVGSRDGTISEDTIRASLISAVSKLRMRKEEMDGA 249
QY 241 QAEINALKRTEEDLKKGHQLKEEMVTRLDQVAEVDKNIELKKKDEELSSALEKMNQS 300
DB 250 QAEINALKRTEEDLKKGHQLKEEMVTRLDQVAEVDKNIELKKKDEELSSALEKMNQS 309
QY 301 ENNDIDEVIITAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
DB 310 ENNDIDEVIITAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 369
QY 361 QFQRLALMQARKTAGLSLDLY 381
DB 370 QFQRLALMQARKTAGLSLDLY 390
```

RESULT 3

```
Q918G8 PRELIMINARY; PRT; 392 AA.
AC Q918G8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Tumor susceptibility protein 101.
OS Chelonia mydas caranigra (Green sea-turtle).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Chelonioidae; Cheloniidae; Chelonia.
OX NCBI_TaxID=8469;
RN [1]
RP SEQUENCE FROM N.A.
RC Yu Q., Lu Y., Nerurkar V.R., Yanagihara R.;
```

RT "Studies on the turtle tumor susceptibility gene, TSG; full-length
RT cDNA sequence, genomic structure analysis and role in green turtle
RT fibropapilloma.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF279276; AAF87776.1; -
DR GO; GO:0004840; Fubiquitin conjugating enzyme activity; IEA.

DR GO; GO:0006512; Ubiquitin cycle; IEA.

DR InterPro; IPR008883; Tsg101.

DR InterPro; IPR000608; Ubq_conjugat.

DR Pfam; PF05743; Tsg101; 1.

DR SMART; SM00212; UBCc; 1.

SQ SEQUENCE 392 AA; 44314 MW; 25D98A5116CF8EA CRC64;

Query Match 90.9%; Score 1813.5; DB 13; Length 392;
Best Local Similarity 89.8%; Pred. No. 1.2e-112;
Matches 343; Conservative 20; Mismatches 18; Indels 1; Gaps 1;

QY 1 MMSKYRDIIVRQTVNIVAMYKDLKPLVDSYVFNDSGSRRLVNLGTIPVYRGNIYNI 60

Db 11 MLAKYKRDITVQETTSVITQYKDLKPLVDVAVFNDSGSRDLMSLTGTPVYRGNTYNI 70

QY 61 PICLWLLDTPVNPPTCFVKPTSSMTIKTKGKVDANGKIYLPYLHDKHPRSELLELIQI 120

Db 71 PICLWLLDTPVNPPTCFVKPTSSMTIKTKGKVDANGKIYLPYLHDKHPRQSDLIQI 130

QY 121 MIVIFGEPPVSRPVSASYPYATGPNTSYMPGMSGISAYPSGYPNPSGYPGCP 180

Db 131 MIVIFGEPPVSRPRTISFQYQATGPNTSYMPGMSGISAYPSGYPNPSGYPGYP 190

QY 181 YPPAGFPATTSQ-YPSQPPVTVGSRDGTISEDTIRASLISAVSKLRWRKEMDG 239

Db 191 YPPGFPATTSQOHTSPVTVGSRDGTISEDTIRASLISAVSKLRWRKEMDR 250

QY 240 AQAEINAKRTBEDLKKHQKLEEMVTRLDQEAIEVDKNIELKKDEELSSALEKWNQ 299

Db 251 AQAEINAKRTBEDLKKHQKLEEMVTRLDQEAIEVDKNIELKKDEELSSALEKWNQ 310

QY 300 SENNDIVIIITAPYKQILNLYAEENAIETIFVLGEALRGVITLDLVFKHVLISR 359

Db 311 SENNDIVIIITAPYKQILNLYAEENAIETIFVLGEALRGVITLDLVFKHVLISR 370

QY 360 KQFQRLALMOKARKTAGLSLDY 381

Db 371 KQFQRLALMOKARKTAGLSLDY 392

RESULT 4

Q9VVA7 PRELIMINARY; PRT; 408 AA.

AC Q9VVA7; Q9GSB5;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DE CG9712 protein (Tumor suppressor protein 101)

GN TSG101 OR CG9712.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

EX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
Ibegwam T.C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuncio J.,
Paclet J., Paragas V., Park S., Patel S., Pfeiffer B.,
Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
"Sequencing of *Drosophila melanogaster* genome.";
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
Ashburner M., Gelbart W.M., Rubin G.M., Murgall C.J., Lewis S.E.;
"Annotation of *Drosophila melanogaster* genome.";
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE FROM N.A.
FlyBase;
Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE FROM N.A.
Laurencon A., Burtis K.C., Hawley S.;
"TSG101 fly homolog.";
Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AE003526; AAF49406.2; -
DR EMBL; AF315343; AAG29584.1; -
DR FlyBase; FBGN0036666; TSG101.

```
DR GO; GO:0004840; F-ubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; P-ubiquitin cycle; IEA.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR InterPro; IPR008883; Tsg101.
DR InterPro; IPR000608; UBQ conjugat.
DR Pfam; PF05743; Tsg101; 1.
DR SMART; SM00212; UBCC; 1.
SQ SEQUENCE 408 AA; 45190 MW; 7C48E20DEB29A19 CRC64;

Query Match 48.3%; Score 964; DB 5; Length 408;
Best Local Similarity 49.3%; Pred. No. 3.8e-56;
Matches 199; Conservative 66; Mismatches 103; Indels 36; Gaps 9;

QY 2 MSKYKRDLTVRQTVNVVIAMKDKLPVLDYVNDGSSRELNVLTGTPVRYRGNITNIP 61
Db :|||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
13 LSKYKVAATKQDVVDVVTGFRSLTYDLQRFVNDGSSKELFTTQGTIPVYVKNNTYIP 72
QY 62 ICILWLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDKHPRSELLELIQIM 121
Db :|||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
73 ICILWLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDKHPRSELLELIQIM 132
QY 122 IVIFGEPPPVFSRP--TVSASYPPTATGPPNTSYM--PGMPGISA---YP-----S 167
Db :|||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
133 IVIFGDHPVYSKEQIAAPY-----PTNSYMPQPGAPGSGNSFLPYPTAGGAGGS 184
QY 168 GYPNPGYGCYP--PAGP-----YPA-----TTSSQVSPPTVTVGPRSDGTISE 214
Db :|||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
185 NFPPYPTGSNVGPYPPTPAGPAGSGYPAPVNFPTQTAGGYPPAAGYNPNPSSGTGITE 244
QY 215 DTIRASLISAVSDKLWRMKEEMDGAQALNALKRTEBDLKGKHKLEEMVTRLDQVAE 274
Db :|||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
245 EHKASITISAIDKLRVRQVEKVNQYQAEITLNRKQELLEGSAKIDAIERLEREHID 304
QY 275 VDKNIELLKKKDELSALEKMNQSENNDIDEVIIITAPLYKQILNLYAEANAIEDTIF 334
Db :|||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
305 MQKNISILKQEKQELKALEDESAEAINP--DEAVTTTAPLYRQLLNAYADEAATEDAIY 363
QY 335 YLGEALRRGVLDLDFLKHVRLLSRKQFQLRALKMOKARTAGLS 378
Db :|||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
364 YLGEALRRGVLDLDFLKHVRLLSRKQFQLRALKMOKARTAGLS 407

RESULT 5
O76258 PRELIMINARY; PRT; 425 AA.
ID O76258
AC O76258;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein C09G12.9.
GN C09G12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peioderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans; a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Le T.T., Harper M.;
RT "The sequence of C. elegans cosmid C09G12.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
```



```
DR InterPro: IPR000608; UBQ_conjugat.
DR Pfam: PF05743; Tsg101; 1.
DR SMART: SM00212; UBCC; 1.
SQ SEQUENCE 249 AA; 27081 MW; 4CAF5F66D029CBB4 CRC64;

Query Match      27.1%; Score 541.5; DB 5; Length 249;
Best Local Similarity 46.5%; Pred. No. 2.5e-28;
Matches 114; Conservative 33; Mismatches 63; Indels 35; Gaps 8;

QY 2 MSKIKYRDLTVRQTVNVIAMKDKPVLDSYFVNDGSSRELVLNLTGTPVRYRGNINYIP 61
Db :||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
13 LSKYKYVAATKKDQVDDVTSFRSLTDLQRFVNDGSSKELFTIQGTIPVYKNNTYIP 72
QY 62 ICLWLDITVYNPFCFVPTSSMTTKTKHVDANGKIYLPYLDHWKHPRSELLELIQIM 121
Db :||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
73 ICIWLDITHPQAPMCFVFKPTPTMQIKVSMYVDHNGKVLYLPYLDHWQPHSSDLLSLIQW 132

QY 122 IVIFGEPPVFSRP--TVSASYPPTATGPNTSYM--PQMPSGISA---YP-----S 167
Db :||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
133 IVTFGDHPVYKPKKEQIAAPY-----PINSYMPQFAGGNSFLPYPTAGGAGGS 184

QY 168 GYPPNPSPGPGCFYP--PAGP-----YPA-----TTSSQYPSQPPVTVGFSRDTGISE 214
Db :||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 NFPPTYGTSNGVGPYPTPAGPAGGYPAYPNFIOPTAGGYPPAAGYNFSFSTGITE 244

QY 215 DTIRA 219
Db : : : : :

QY 245 EHIKA 249
Db : : : : :

RESULT 7
Q7T2M2 PRELIMINARY; PRT; 114 AA.
AC Q7T2M2;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Tumor susceptibility protein 101 (Fragment).
OS Ameiurus nebulosus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ameiurus.
OX NCBI_TaxID=27778;
RP SEQUENCE FROM N.A.
RA Busch C.R., Heath D.D., Hubberstey A.;
RT "The impact of PAHs on the apoptotic mechanisms in the brown bullhead
  (Ameiurus nebulosus).";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY219035; AAP69601.1; -.
FT NON_TER 114
FT NON_TER 114
SQ SEQUENCE 114 AA; 13033 MW; 85FCE752C6BB4387 CRC64;

Query Match      25.3%; Score 505; DB 13; Length 114;
Best Local Similarity 75.4%; Pred. No. 2.6e-26;
Matches 86; Conservative 17; Mismatches 11; Indels 0; Gaps 0;

QY 11 TVRQTVNVIAMKDKPVLDSYFVNDGSSRELVLNLTGTPVRYRGNINYIPICLWLLDTY 70
Db :||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 TVRQTVNVIAMKDKPVLDSYFVNDGSSRELVLNLTGTPVRYRGNINYIPICLWLLDTY 60

QY 71 PYNPPICFVKPTSSMTTKTKHVDANGKIYLPYLDHWKHPRSELLELIQIMVI 124
Db :||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 PYNPPICFVKPTSSMTTKTKHVDANGKIYLPYLDHWKHPRSELLELIQIMVI 114

RESULT 8
Q9LHG8 PRELIMINARY; PRT; 398 AA.
AC Q9LHG8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
RT "Full-length messenger RNA sequences greatly improve genome
```

```
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Mouse and human tumor susceptibility gene-like protein (Hypothetical
  protein).
DN T2E22_28 OR T2E22_129/AT3G12400 OR AT3G12400.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
  eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
  Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
  Sequence features of the regions of 4,251,695 bp covered by ninety Pl.
  TAC and BAC clones.";
RL DNA Res. 7:217-221 (2000).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansonge W., Unseid M.,
  Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
  Delsen M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
  De Simone V., Choisme N., Artiguenave F., Robert C., Brottier P.,
  Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
  Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
  Würlbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
  Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
  Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
  Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
  Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
  Navarro P., Collado C., Perez-Perez A., Purnelle B., Masuy D.,
  Cooke R., Laudie M., Berger-Lilauro C., Ottenwaelder B., Duchemin D.,
  de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
  Monfort A., Argioui A., Flores M., Lignori R., Vitale D.,
  Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
  Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
  Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
  Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
  Pai G., Militscher J., Sellers P., Gill J.E., Feldlyum T.V.,
  Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
  Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
  Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
  Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
  Nakayama S., Nakazaki N., Shimp S., Takeuchi C., Wada T.,
  Watanabe A., Yamada M., Yasuda M., Tabata S.;
  "Sequence and analysis of chromosome 3 of the plant Arabidopsis
  thaliana.";
RL Nature 408:820-822 (2000).
[4]
RP SEQUENCE FROM N.A.
RA Yamada K., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
  Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
  Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
  Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
  Kosuma E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
  Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
  Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
  Theologis A.;
  "Full length cDNA of gene T2E22_129/At3g12400 (GI:15230478).";
  Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
  Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
  "Full-length messenger RNA sequences greatly improve genome
```

```

RT annotation.";
RL Genome Biol. 0:0-0(2002).
RN [6]
RP SEQUENCE FROM N.A.
RA Feldmann K.;
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RT "Full-length cDNA from Arabidopsis thaliana."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C.,
RA Wu H.C., Yu G., Yuan S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Lam B., Lin J., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AP020477; BAB03147.1; -.
DR EMBL; AC069474; AAG51025.1; -.
DR EMBL; AY056283; AAL07132.1; -.
DR EMBL; AY087782; AAM65318.1; -.
DR EMBL; AY133779; AAM91713.1; -.
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA.
DR InterPro; IPR008883; Tsg101.
DR Pfam; PF05743; Tsg101; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 398 AA; 44715 MW; 659956F9784FDC20 CRC64;

Query Match 22.0%; Score 440; DB 10; Length 398;
Best Local Similarity 31.6%; Pred. No. 2.5e-21;
Matches 117; Conservative 72; Mismatches 147; Indels 34; Gaps 12;

QY 12 VQ--TVNVIAMKDKLPVLSYVFNDSGSSRELNLNLTCTIPVRYGNINYNIPICLWLDTY 70
DB 37 IQHLLNLISSYPSLEPTASFMHNDGSRVNLNQADGTIPMPFHGVYINPIVILLESY 96
QY 71 PYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLDHDKHPRSELLELIQIMVIFGEEP 129
DB 97 PRHPCCVNVNPTADMIKRPFAHVTPSGLVSLPYLQNVVYVPSNLDVLSLSAFAFARDP 156
QY 130 PVFSRPTVSASYPPTATGNPT--SYMPGMPGISAPSGYPNPSPGCPFPAGPY 187
DB 157 FLXSR---RRPQP--PPSPPTVYDSSLRPPSADQSLRPPFPFPSPYG-----GGVS 203
QY 188 PATTSSOYPSQPPVTVTGPSRDGTISE--DTIRASLISAVSDKLRWRMKEEMDGAQLN 245
DB 204 RVQVQVHHQOQSDDAEVEFKENAINKVMVHSDLV--MRRAREAEAEELLSQA--- 258
QY 246 ALKTEEDLKGGHQLBEMVTRLDQEAVDNKNIELLKXKDEELSGALEKMNENSEN--- 302
DB 259 GLKREDELNIG---LKEMV---EEKETLPQQLIISMNTDILDSVWVRENQGTKNLVD 311
QY 303 NDIDEVITPAPLYKOILNLYAEENATDTTFYLGAEALRGVIDLVFLKHVRLLSRQF 362
DB 312 LDVDNAFECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVFPDQYLRNVRLLSREQ 371
QY 363 QLRALMOKAR 372
DB 372 FHRATGSKVR 381

RESULT 9
Q8IX04
ID Q8IX04 PRELIMINARY; PRT; 379 AA.
AC Q8IX04;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ubiquitin-conjugating enzyme E2 variant.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22314737; PubMed=12427560;
RA Kloor M., Bork P., Duwe A., Klaes R., von Knebel Doeberitz M.,
RA Ridder R.;
RT "Identification and characterization of UBV3, a human cDNA with
RT similarities to inactive E2 ubiquitin-conjugating enzymes."
RL Biochim. Biophys. Acta 1579:219-224(2002).
DR EMBL; AF503350; AAN32950.1; -.
DR GO; GO:0004459; F:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0014691; F:oxidoreductase activity; IEA.
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA.
DR InterPro; IPR001236; ldh.
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00056; ldh; 1.
DR Pfam; PF02866; ldh; 1.
DR PRINTS; PR00086; LLDHGRGNASE.
SQ SEQUENCE 379 AA; 42267 MW; 100BCD1FA290E860 CRC64;

Query Match 20.8%; Score 415; DB 4; Length 379;
Best Local Similarity 54.9%; Pred. No. 1.1e-19;
Matches 73; Conservative 26; Mismatches 34; Indels 0; Gaps 0;

QY 1 MMSKYKVDLTVRQTVNVIAMKDKLPVLSYVFNDSGSSRELNLNLTCTIPVRYGNINYNI 60
DB 11 LLGKYKFRDUTVSELNRNVNVPFPHKYSMDTIYVFKSSQKDLNFTGTIPVYQNTYNI 70
QY 61 PICLWLDTYVNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLDHDKHPRSELLELIQI 120
DB 71 PIRFWILDSHPFAPPICFLKPTANMGLVKGHVDAQGRILYLPYLNQMSHPKSVIVGLIKE 130
QY 121 MIVIFGEPPVFS 133
DB 131 MIAKFOBELEPYS 143

RESULT 10
Q08761
ID Q08761 PRELIMINARY; PRT; 174 AA.
AC Q08761;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Signaling molecule.
GN ATP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=Thymus;
RA Burbello P.D.;
RT "Role of ATP in signal transduction."
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U97571; AAB57699.1; -.
DR MGD; MGI:1860490; Attp.
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA.
DR InterPro; IPR000608; UBQ_conjugat.
DR SMART; SM00212; UBCc; 1.
SQ SEQUENCE 174 AA; 20499 MW; 81BEE363DD81900B CRC64;

Query Match 20.6%; Score 412; DB 11; Length 174;

```

	Best Local Similarity	53.3%;	Pred.	No.	6.6e-20;	Gaps	0;		
	Matches	72;	Conservative	27;	Mismatches	36;	Indels	0;	
QY	1	MMSKYKRDLDLTVRQTNNVIAMYSKDLPVDSYVFNDGSSRELNLVTGTIPVRYRGNTYNI	60	:	: :: :::	:::	: :: :::	::	
Dd	11	LLGKYFRDLTVEELKNVSFPHFRYSVDIVYFKDTSQDLLNFTGTIPVMYQKTNYI	70	:	: :: :::	:::	: :: :::	::	
QY	61	PICLWLLDTYPNPPIPCVTKTSMTTIKGHVDAANGKIYILPYLHDWKHPSELLELIQI	120	:	: :: :::	:::	: :: :::	::	
Dd	71	PIRFWLSDHFPAPPICFLPKPTANMEISVGKHVDAGRIYLYPLQNWSHPKSAIVGLIKE	130	:	: :: :::	:::	: :: :::	::	
QY	121	MIVIFGEPPPVESRP	135	:	: :: :::	:::	: :: :::	::	
Dd	131	MIARFOQLPAYSIP	145	:	: :: :::	:::	: :: :::	::	
RESULT	11								
QYTQD3	ID	Q7TQD3	PRELIMINARY;	PRT;	249 AA.				
AC	Q7TQD3,								
DT	01-OCT-2003	(TrEMBLrel. 25, Created)							
DT	01-OCT-2003	(TrEMBurel. 25, Last sequence update)							
DT	01-OCT-2003	(TrEMBurel. 25, Last annotation update)							
DE	Hypothetical protein.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_TaxID=10090;								
[1]	SEQUENCE FROM N.A.								
RP	STRAIN=C57BL/6; TISSUE=Brain;								
RC	MEDLINE=22388257; PubMed=12477932;								
RX	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Boone T., Max S.I., Wang J.J., Hsieh F., Hopkins R.F., Jordan H., Moore T., Kach S.R., Rubin G.M., Hong L., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshikiyuki S., Carninci P., Prange C., Raha S.S., Lequellano N.A., Peters G.J.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Bullyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Wahby J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences"; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002). [2]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=C57BL/6; TISSUE=Brain;								
RX	Strausberg R.; Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.								
RL	EMBL; BC054796; AAHS4796.1; -.								
DR	Hypothetical protein.								
XQ	SEQUENCE 249 AA; 27953 MW; 5CD3B6724D374C89 CRC64;								
Query Match		20.4%;	Score	408;	DB 11;	Length	249;		
Best Local Similarity		35.5%;	Pred.	No.	1.9e-19;				
Matches	94;	Conservative	47;	Mismatches	90;	Indels	34;	Gaps	7;
QY	1	MMSKYKRDLDLTVRQTNNVIAMYSKDLPVDSYVFNDGSSRELNLVTGTIPVRYRGNTYNI	60	:	: :: :::	:::	: :: :::	::	
Dd	11	LLGKYFRDLTVEE-LNVSVSFFPHFRYSVDIVYFKDTSQDLLNFTGTIPVMYQKTNYI	69	:	: :: :::	:::	: :: :::	::	
QY	61	PICLWLLDTYPNPPIPCVTKTSMTTIKGHVDAANGKIYILPYLHDWKHPSELLELIQI	120	:	: :: :::	:::	: :: :::	::	
Dd	70	PIRFWLSDHFPAPPICFLPKPTANMEISVGKHVDAGRIYLYPLQNWSHPKSAIVGLIKE	129	:	: :: :::	:::	: :: :::	::	

```
QY 352 KHVLLSRKQFQRLQALMQARK 373
Db 331 RNVLLSRQFQFHRATAEKVRE 352

RESULT 13
Q96FF5 PRELIMINARY; PRT; 177 AA.
AC Q96FF5
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011011; AH01011.1; -.
DR GO; GO:0004840; Fubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; Fubiquitin cycle; IEA.
DR InterPro; IPR000608; UBQ_conjugat.
DR SMART; SM00212; UBCC; 1.
KW Hypothetical protein.
SQ SEQUENCE 177 AA; 19552 MW; E733E710F8D0069A CRC64;

Query Match 18.1%; Score 361; DB 4; Length 177;
Best Local Similarity 59.0%; Pred. No. 1.7e-16;
Matches 62; Conservative 21; Mismatches 22; Indels 0; Gaps 0;

QY 29 LDSYFVNDGSSRELNLGTIPVYRGNIYNIPICLWLLDTPYNPPICFVKPTSSMTIK 88
Db 1 MDTYVFKDSQKLLNFTGTIPVMYQNTYNIPIREWILDSHPFAPPICFLKPTANMGIL 60

QY 89 TGHVDANGKIYLYLHDWKHPRSELLELIQIMIVIFGEPPVPS 133
Db 61 VGRHVDAGRIYLYLQNWSHPKSVIVGLIKEMIAKFQBELPMYS 105

RESULT 14
Q9NUX7 PRELIMINARY; PRT; 341 AA.
AC Q9NUX7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FJ11068 (EC 1.1.1.27) (L-lactate dehydrogenase)
DE (LDH).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsumura H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: (S)-LACTATE + NAD(+) = PYRUVATE + NADH.
CC -!- PATHWAY: ANAEROBIC GLYCOLYSIS; FINAL STEP.
CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY.
DR EMBL; AK001930; BAA91985.1; -.
DR HSSP; P00336; 5LDH.
DR GO; GO:0004459; F:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.
```

```
DR GO; GO:0006096; P:glycolysis; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA.
DR InterPro; IPR001236; Ldh.
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00056; Ldh; 1.
DR Pfam; PF02866; Ldh.C; 1.
DR PRINTS; PR00086; LLDHGRGNASE.
DR SMART; SM00212; UBCC; 1.
KW Hypothetical protein; Glycolysis; NAD; Oxidoreductase.
SQ SEQUENCE 341 AA; 37587 MW; 7D5CD535296165FA CRC64;

Query Match 18.1%; Score 361; DB 4; Length 341;
Best Local Similarity 59.0%; Pred. No. 3.7e-16;
Matches 62; Conservative 21; Mismatches 22; Indels 0; Gaps 0;

QY 29 LDSYFVNDGSSRELNLGTIPVYRGNIYNIPICLWLLDTPYNPPICFVKPTSSMTIK 88
Db 1 MDTYVFKDSQKLLNFTGTIPVMYQNTYNIPIREWILDSHPFAPPICFLKPTANMGIL 60

QY 89 TGHVDANGKIYLYLHDWKHPRSELLELIQIMIVIFGEPPVPS 133
Db 61 VGRHVDAGRIYLYLQNWSHPKSVIVGLIKEMIAKFQBELPMYS 105

RESULT 15
Q873M7 PRELIMINARY; PRT; 378 AA.
AC Q873M7
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Vacuolar protein sorting 23.
GN VPS23.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA Blanchin-Roland S., Gaillardin C.;
RT "New genes involved in genetic control of extracellular protease
RT synthesis in Yarrowia lipolytica.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Blanchin-Roland S.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ509167; CAD48920.1; -.
SQ SEQUENCE 378 AA; 41788 MW; C5498CACC5ED3669 CRC64;

Query Match 14.9%; Score 297; DB 3; Length 378;
Best Local Similarity 27.5%; Pred. No. 7.6e-12;
Matches 104; Conservative 48; Mismatches 140; Indels 86; Gaps 11;
```

```
QY 85 MTIKTKGHVDANGKIYLYLHDW-KHPR-SELLELIQIMIVIFGEPPVSRPTVSASP 142
Db 1 MMLSPGNYVDNNGRECHYPYISEWGSDDPQNTNLATFLVLSDFISKEPPVSRPTVEYGP 60

QY 143 P-----YTATGPPNTSYMPGMP-----SGISAYP----- 166
Db 61 PTPSPONDQOQTHQMQRPPPLAQSGMPFIQFVOTGQHAVPPVFNRTQCHIQQOQPMRTGPV 120

QY 167 -----SGVPPNFGVYP-----GCPYPAGPYPATTSQYP---SQPPVTTVGP- 206
Db 121 DLGSSSTHASEAVPPLDPKPPQWQNGAVVQPPAQYAYAHQHPLOASHHPQARGGPL 180

QY 207 -----SRDGTISEDTIRASLISA-----VSKLWRMKEMMDGAQAE 243
Db 181 SQQFTPHRGRASRTDIMDMDTAKSDEPAPPKPNPERMKALDLDLSQLKEADAIOAN 240

QY 244 LNALKRTEDLKKGHQKLEEWVTR----LDQEAFAVDKNIELIKKKDEELSALEKMEQ 299
```

```

Db 241 VOADDAQIEGLWKLSSLEAGVSSEVLLRLRLEAQADKNNKILSDRMQOARVITQ-ARA 299
QY 300 SENNDIDEVITPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVVIDLDVFLKHVRLLSR 359
Db 300 CEIPDIDSACVCAENVFNQYLDLTAGEQAIDTIIYALSALDREKIVTEFFPKHVRNLAR 359
QY 360 KQFQRLALMOKARKTAGL 377
Db 360 EKFIKVIATIDKIAVGAGL 377

RESULT 16
ID P87279 PRELIMINARY; PRT; 385 AA.
AC P87279;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE STP22P.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=XG1;
RA Li Y., Kane T., Tipper C., Spatrick P., Jenness D.D.;
RL "The complete sequence of STP22 gene.";
DR EMBL: AF004731; AAB62820.1; -
DR GO: GO:0005768; C: endosome; IDA.
DR GO: GO:0005515; F: protein binding; IDA.
DR GO: GO:0006612; P: protein-membrane targeting; IMP.
DR GO: GO:0006623; P: protein-vacuolar targeting; IMP.
DR InterPro: IPR008883; Tsg101.
DR InterPro: IPR000608; UBO conjugat.
DR Pfam: PF05743; Tsg101; 1.
DR SMART: SM00212; UBCC; 1.
DR PROSITE: PS0127; UBIQUITIN CONJUGAT 2; 1.
SQ SEQUENCE 385 AA; 43330 MW; FEDE3BE79F063BCE CRC64;

Query Match 14.7%; Score 293.5; DB 3; Length 385;
Best Local Similarity 26.0%; Pred. No. 1.3e-11;
Matches 102; Conservative 74; Mismatches 164; Indels 53; Gaps 16;

QY 7 YRD--LTVQTVNVIAMKDLKPLVDLSYVFNDSGSRRLVNLGTIPVRYRGNL-YNIPIC 63
Db 25 YNDGRTTFHDSLALLDNFHSRLRPRTRVTHSDGTPQLLSIVGTISTGEDSSPHSIV 84
QY 64 LWLDDTYNPPIC-----FVKPTSSMTIKGKHVDANGKIYLPYLDHDKHPRSELLEL 117
Db 85 MWVFSMPYVKPPFISINLENDFMNTISSLSPIQEIYDSNGWIALPILHCWDPAAANLIM 144
QY 118 IQIMVIFGEPFPPFRTVSASVPPYAT---GPPNTSYMGPSGSIAYSPGVPNPS 174
Db 145 VQELMSLL-HPPQDQAB-----SLPPKNTQLQEQNTPLPPPKS-----PHLKPPLP- 194
QY 175 GYPGCPYPPAGPYPATTSQYSPQPPVTTPGFSRD-----GTISEDTIRASLISAVSD 227
Db 195 -----PPPPQPPASNALDMDN-DTISPTNHEMLQNLQTVNVNLYRED-VDYVAD 245
QY 228 KL--RWRMKMDGQAELNALKRTEEDLKGHQKLEEMVTRLDQEVAVDNKLTLLKKK 285
Db 246 KILTRQTMQESIARFHFIIAIDKNH--LRAVEQAIEQTMESLNAQIDVLTAN----RAK 299
QY 286 DEELSSALEKMNQENNDIDIVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVI 345
Db 300 VQFSST-----SHVDDSEVNSIAVAKTDGLNQLNVAQDYALTDTIECLSRMLHRTGI 354
QY 346 DLDFVLKHLVRLSRKQFQRLALMOKARKTAGLS 378
Db 355 PLDTFVKQRELARQQFLVRWHIQ--RITSPLS 385
```

```

RESULT 17
QYBU96 PRELIMINARY; PRT; 111 AA.
AC QYBU96;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Inferred: signaling molecule ATPP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RL "Analysis of the mouse transcriptome based on functional annotation of
  60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL: AK086750; BAC39736.1; -
DR GO: GO:0004840; F: ubiquitin conjugating enzyme activity; IEA.
DR GO: GO:0006512; P: ubiquitin cycle; IEA.
DR InterPro: IPR000608; UBO conjugat.
SQ SEQUENCE 111 AA; 12909 MW; 78EACC8158EB92B3 CRC64;

Query Match 14.2%; Score 283; DB 11; Length 111;
Best Local Similarity 57.1%; Pred. No. 1.4e-11;
Matches 48; Conservative 17; Mismatches 19; Indels 0; Gaps 0;

QY 53 YRGMYNIPICLLLDITVYPNPPTICFVKPTSSMTIKGKHVDANGKIYLPYLDHDKHPRS 112
Db 2 YQKTYNIPIRFWILDSHPFPAPICFLAPTANMEISVGKHVDAGKIYLPYLDHDKHPRS 61
QY 113 ELLELIQIMVIFGEEPPVFSRPT 136
Db 62 AIVGLIKEMIAKFOELFLYSIPS 85

RESULT 18
QYBU96 PRELIMINARY; PRT; 296 AA.
AC QYBU96;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN STP22.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92245758; PubMed=1574926;
RA Bolle P.A., Gilliquet V., Berben G., Dumont J., Hilger F.;
RL "Yeast sequencing reports: The complete sequence of K3B, a 7.9 kb
  Fragment between PGK1 on chromosome III, reveals the presence of seven
  open reading frames.";
RL Yeast 8:205-213 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92244356; PubMed=1574125;
RA Oliver S.G., der Aart Q.J., Agostoni-Carbone M.L., Aigle M.,
RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.,
RA Benit P.;
RL "The complete DNA sequence of yeast chromosome III [see comments].";
RL Nature 357:38-46 (1992).
```



```

QY 228 KLRWR---MKEMDGAQELNALKTEEDLKGKHOKLEMYRTRLDQVAVEDKNIELKK 284
Db 246 KILTRQTVNQE-----SIARFHEIIADKNHLRAVEQAIQTMSHNAQIDVNLN 295
QY 285 K 285
Db 296 K 296

RESULT 19
P78998 PRELIMINARY; PRT; 376 AA.
AC P78998;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN LGVCLUNSC.
OS Saccharomyces pastorianus (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=27292;
RN [1]
RP SEQUENCE FROM N.A.
RA Andersen T., Nilsson-Tillgren T.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z86109; CAB06793.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA.
DR InterPro; IPR008195; Ribosomal_L34E.
DR InterPro; IPR008883; Tsg101.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF05743; Tsg101; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS01145; RIBOSOMAL_L34E; 1.
DR PROSITE; PS01027; UBIQUITIN_CONJUGAT_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 41542 MW; 837F91DA97BF0A0B CRC64;

Query Match 8.4%; Score 167.5; DB 3; Length 376;
Best Local Similarity 19.3%; Pred. No. 0.0031;
Matches 64; Conservative 59; Mismatches 127; Indels 81; Gaps 9;

QY 11 TVRQTWNVIAMVKDKPVLDVSFVNDGSSRELVLNLTGTIPVRYRGNINIPICLWLLDTY 70
Db 31 TFHDSLALLDKFHLRPRTRVTFHSDGSPQLLSYGTIV----GD--SLPLIMWIPSLY 83
QY 71 PYNPPI-----CFVKPTSSMTIKTKGHVDANGKIYLPYLDHDKHPRSELLELIQIMIVI 124
Db 84 PVKPPFISIDLETDFVSAISSSLPQVAYIDSGRVALPILDHWNATWGLIIVVQELMSL 143
QY 125 FGEEPPVFSRPTVSASYPYATGPNNTSYMPGMPSGISAYPSGYPNPSPGPGCPYPPA 184
Db 144 LSE-----PSRDRVPSLPKPKFDAMHS-----PLKHTVPFLPPK 176
QY 185 GPYPATTSQYSPQPPVTTVGSRDGTISEDITIRASLISAV-----SDKLRWEMKEM 237
Db 177 AKFPHV---QPPLQPPPPPPQPPS-----SAVLDMDMDNTDLSPTNHHWML 218
QY 238 DGAQAEALNALK-----TEEDLKGKHOKLEMYRTRLDQVAVEDKNIEL 281
Db 219 QNLQSVNLNELYEDVHYVADKILTRCTIMQDSVARFHEWAVDKTRLQAVEQTIEQTWHT 278
QY 282 LKKKDEELSSALEKMEQSENNDIETIPT 312
Db 279 LNAQIEVLTAERATVQEFSSSTPFDEEDVDIT 309
```

```

RESULT 20
Q9NTQ8 PRELIMINARY; PRT; 148 AA.
AC Q9NTQ8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN DKFZP564J157.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansorge W., Bloecher M., Bloecher H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435 (2001).
DR EMBL; AL122042; CAB59179.2; -.
KW Hypothetical protein.
SQ SEQUENCE 148 AA; 15375 MW; 796D11BA16A0591E CRC64;

Query Match 7.6%; Score 151; DB 4; Length 148;
Best Local Similarity 28.7%; Pred. No. 0.012;
Matches 39; Conservative 15; Mismatches 32; Indels 50; Gaps 6;

QY 135 PTVSASYPYATGPNNTSYMPGMPSGISAYPSGYPNP-----SGYPGC---PYPPAGPY 187
Db 30 PPINPPPPGCPPPP-----PGAHGPNPAPFSGGPHVPQPGYGGCQLPGYPPYPP 83
QY 188 PATTSSQYSPQPPVTTVGSRDGTISEDITIRASLISAVSDKLRWEMKEMDGAQAEALNAL 247
Db 84 PA-----PGIPPNPLAPGMVGP-----AVVD----- 106
QY 248 KRTEEDLKGKHOKLEE 263
Db 107 KRMQRKMKKHAHKMKH 122

RESULT 21
Q921F1 PRELIMINARY; PRT; 503 AA.
AC Q921F1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Similar to annexin A11.
GN ANXALL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- DOMAIN: A PAIR OF ANNEXIN REPEATS MAY FORM ONE BINDING SITE FOR
CC CALCIUM AND PHOSPHOLIPID (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
DR EMBL; BC012875; AAH12875.1; -.
DR MGD; MGI:108481; Anxall.
DR GO; GO:0005635; C:nuclear membrane; ISS.
DR GO; GO:0005654; C:nucleoplasm; ISS.
DR GO; GO:0005515; F:protein binding; ISS.
DR InterPro; IPR001464; Annexin.
```

DR Pfam; PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXIN.
DR ProDom; PD000143; Annexin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXIN; 4.
KW Annexin; Calcium; Calcium-binding; Calcium/phospholipid-binding;
KW Repeat.
SQ SEQUENCE 503 AA; 54079 MW; 33F3471BE21A0D32 CRC64;

Query Match 7.4%; Score 147; DB 11; Length 503;
Best Local Similarity 26.3%; Pred. No. 0.1;
Matches 77; Conservative 29; Mismatches 95; Indels 92; Gaps 16;

QY 125 FGEPVFRPTVSAYPPVTATGPPNTSYMPGSGISAYPSGYPNPSGYPGCPVPPA 104
DB 89 FGQPPA-QQP-----PPGMPPPGNGPPGMPG-----YPA-YGAP--VFGQMPPT 135
QY 185 GPYPATSSQYSPQPPVTG-----PSRDGTISEDTIRASL-----I 222
DB 136 GQQP--PGAYPQPQPMTPGQSPMPPPGQPPVPSYSGSSTITPAVPPAQFGNGRTI 192
QY 223 SAVS-----DKLRWKEEMDGAQAEINALKRTEEDLKGKHKLEEMVTRLDQVAE 274
DB 193 TAASGFDPDLDAEVLKAMKMGFGTDEQAIDCLGSRN--KQRQQLLSFKTAYGK-- 246
QY 275 VDKNIELKKKDELSALEKMKENQSENNDIDEVLIPTAPLYKQILNLYAEANAIEDTIF 334
DB 247 -----DLIKDKLSLGNFE-----KTLALM--KTFVLFDVY 277
QY 335 YLGEALRRGVID----LDVFL----KRVLLSRKQFQRLALMQKARKTAGLSD 379
DB 278 EIKEAIKAGTDEACLIIEIPASRNEHIELSR---AYKTEFQKLTLEAIRSD 327

RESULT 22

Q9NZ81 PRELIMINARY; PRT; 148 AA.
ID Q9NZ81
AC Q9NZ81;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Uncharacterized bone marrow protein BM041 (DKFZp564J157 protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Zhao M., Gu J., Li N., Peng Y., Han Z., Chen Z.;
RT "A novel gene expressed in human bone marrow."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217517; AAF67628.1; -
DR EMBL; BC016064; AAH16064.1; -
DR EMBL; BC014257; AAH14257.1; -
SQ SEQUENCE 148 AA; 15385 MW; 56BFE6A15935A2E CRC64;

Query Match 7.3%; Score 146; DB 4; Length 148;
Best Local Similarity 27.9%; Pred. No. 0.026;
Matches 38; Conservative 15; Mismatches 33; Indels 50; Gaps 6;

QY 135 PTVSAYPPVTATGPPNTSYMPGSGISAYPSGYPNP---SGYPCG----PYPPAGPY 187
DB 30 PPINPPPPGPCPPP-----PGAPHGNAFPFGPPHPVPPGYPGCPGLGYPYPPYP 83
QY 188 PATSSQYSPQPPVTGTPSRDGTISEDTIRASLISAVSDKLRWKEEMDGAQAEINAL 247
DB 84 PA-----PGIPPNFLAPGMVGP-----AVVD----- 106

QY 248 KRTEEDLKGKHKLE 263
DB 107 KRMQKTKKKAHKQHK 122

RESULT 23
Q8IZL8 PRELIMINARY; PRT; 1130 AA.
ID Q8IZL8
AC Q8IZL8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MNAR.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-22317447; PubMed-12415108;
RA Wong C.W., McNally C., Nickbarg E., Komm B.S., Cheskis B.J.;
RT "Estrogen receptor-interacting protein that modulates its nongenomic
RT activity-crosstalk with Src/Erk phosphorylation cascade."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14783-14788(2002).
DR EMBL; AF547989; AAN41255.1; -
DR InterPro; IPR008938; ARM.
DR InterPro; IPR002965; P rich extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1130 AA; 119623 MW; 96A698015E3DD7CF CRC64;

Query Match 7.3%; Score 145; DB 4; Length 1130;
Best Local Similarity 26.4%; Pred. No. 0.37; Indels 26; Gaps 8;
Matches 51; Conservative 33; Mismatches 83;

QY 122 IVIFGEE-PPVSRPTVSASYPPTATGPPNTSYMPGMPG-----GISAYPSGYPNP 175
DB 788 VVIVPSGLPLPPPPSGATPPPIAPTGPTAS--PPVPAKEPPEELPAAPGFLPPP-- 843
QY 176 YFCGPPYPAGYPATTSQYSPQPPVTGTPSRDGTISEDTIRASLISAVSDKLRWKE 235
DB 844 -----PPPPPPGVPVLPFPQVPEGTGGGPPALEEDLTVININSDDEE-----EE 892
QY 236 EMDGAQAEINALKRTEEDLKGKHKLEEMVTRLDQVAEVDKNIELKKKDELSALEX 295
DB 893 EEEGEEE-EEEEEEEDFEEDDEEYFEEEEEEFEFEFEFEFEFEFEFEFEFEFEFE 947
QY 296 MENQSENNDIDEV 308
DB 948 EEEDEEEEEELEEV 960

RESULT 24
Q9BU60 PRELIMINARY; PRT; 746 AA.
ID Q9BU60
AC Q9BU60;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002875; AAH02875.1; -
DR InterPro; IPR002965; P rich extensn.
DR PRINTS; PR01217; PRICHEXTENS.

[illegible]

```

RESULT 25
Q96FTI
ID ID Q96FTI PRELIMINARY; PRT; 1052 AA.
AC Q96FTI;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
[1]
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010457; AAH10457.1; -.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR002965; P rich extensn.
DR PRINTS; PR01217; PRICEXTENS.
KW Hypothetical protein.
FT
SQ SEQUENCE 1052 AA; 112161 MW; C71A751164C8CIAB CRC64;

```

Query Match	7.2%	Score 143;	DB 4;	Length 1052;
Best Local Similarity	25.9%	Pred.No. 0.46;		
Matches 50;	Conservative 32;	Mismatches 85;	Indels 26;	Gaps 7;
QY	122	IVIFGEE--PPVFSRPTVSASYPPTATGCPPTSYMPGMPSS-----GISAVSGYGPYPNPSG	175	
Db	710	VWVPEGLPLPPPPPSGATPPPTATGCPPTAS--PPVPAKEEPEEELPAPGPIPPPP--	765	
QY	176	YPGCPYPAGYPANTSSQYPSQPVTTVTGFSRDTISDTIRASLISAVSDKILRWKME	235	
Db	766	-----PPPPVPQGVTLPPPLQVPEGTGGGGPPALEEDLTVININSDEE-----EE	813	
QY	236	EMDCAQARLNALKTEEDLKKGHQKLEWTVTRLQEVAEVDKNTELLKKKDEELSSALEK	295	
Db	814	EEEEEEEEEEEEEEEDDFEEEEDEEYFYEEEEEEEEEPPEPFE---EEGELEEE--EE	869	
QY	296	MENQSENNDIDEV	308	
Db	870	EEEEEEEEEEFV	882	

Search completed: July 12, 2004, 08:32:10
Job time : 45.5506 secs

GenCore version 5.1.1.6
 Copyright (c) 1993 - 2004 CompuGen Ltd.
 OM protein - protein search, using sw model
 Run on: July 12, 2004, 08:23:46 ; Search time 50.4336 Seconds
 (without alignments)
 2128.898 Million cell updates/sec
 Title: US-09-804-690-4
 Perfect score: 2002
 Sequence: 1 MVSKYKRDILVREIVNVIT.....FQLRALMOKARKTAGLSDLV 380

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 100 summaries

Database : A.Geneseq 29Jan04: *
 1: geneseqp1980s: *
 2: geneseqp1990s: *
 3: geneseqp2000s: *
 4: geneseqp2001s: *
 5: geneseqp2002s: *
 6: geneseqp2003as: *
 7: geneseqp2003bs: *
 8: geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2002	100.0	380	2	AAW19111 Human tum
2	2002	100.0	390	2	AAW93424 Human NHT
3	2002	100.0	390	5	ABF52191 Human tum
4	2002	100.0	390	6	ABG73787 Human tum
5	2002	100.0	390	7	ADC35176 Novel hum
6	1997	99.8	390	6	AAE34885 Human Tsg
7	1900.5	94.9	381	2	AAW19110 Mouse tum
8	1900.5	94.9	381	2	AAW93425 Mouse tsg
9	1900.5	94.9	381	6	ABG73788 Murine ts
10	1900.5	94.9	381	7	ADC35178 Human Tsg
11	1900.5	94.9	391	6	AAE34884 Mouse Tsg
12	1900.5	94.9	391	7	ADB85228 Mouse tum
13	1404	70.1	307	5	ABP41729 Human ova
14	1117	55.8	237	4	AAE09328 Human int
15	809.5	40.4	331	4	ABBE4607 Drosophil
16	455	22.7	398	3	AAAG06370 Arabidops
17	455	22.7	412	3	AAAG06369 Arabidops
18	434	21.7	90	3	AAAG01689 Human sec
19	427	21.3	379	5	ABP43551 L-lactate
20	424	21.2	340	3	AAAG06371 Arabidops
21	380	19.0	322	3	AAAG13898 Arabidops
22	371	18.5	341	4	AAAB93473 Human pro
23	346	17.3	146	3	AAAB53717 Human col
24	331	16.5	452	4	AAU23116 Novel hum
25	267.5	13.4	288	3	AAAG13899 Arabidops

26	161.5	8.1	168	3	AAAG13900	Aag13900 Arabidops
27	161	8.0	485	3	AAAB57086	Aab57086 Human pro
28	160	8.0	466	6	ABO07215	Abo07215 Human p53
29	160	8.0	466	7	ADD47589	Add47589 Human Pro
30	153	7.6	218	4	ABG18833	Abg18833 Novel hum
31	151	7.5	466	2	AAAR07084	Aar07084 Recombina
32	148.5	7.4	148	3	AAIF86515	Aay86515 Human gen
33	148.5	7.4	148	5	ABB97431	Abb97431 Novel hum
34	148.5	7.4	148	6	ABO53679	Abo53679 Novel hum
35	148.5	7.4	176	4	AAAG75132	Aag75132 Human col
36	148.5	7.4	397	4	AAAM39852	Aam39852 Human pol
37	148.5	7.4	397	4	AAAG81381	Aag81381 Human APF
38	148.5	7.4	397	6	AAAS5415	Aas5415 Human pro
39	148.5	7.4	442	4	AAAM41638	Aam41638 Human pol
40	144.5	7.2	777	5	ABP41839	Abp41839 Human ova
41	144.5	7.2	1130	7	ADC99065	Adc99065 Human KPP
42	142	7.1	505	2	AAAY07117	Aay07117 Lung canc
43	142	7.1	505	6	ABO07216	Abo07216 Human p53
44	142	7.1	905	2	AAAW31186	Aaw31186 Human p16
45	142	7.1	1135	2	AAAW31185	Aaw31185 Human p16
46	142	7.1	1251	7	ADE09332	Ade09332 Novel pro
47	141	7.0	244	3	AAAG40015	Aag40015 Arabidops
48	141	7.0	245	3	AAAG38059	Aag38059 Arabidops
49	141	7.0	245	3	AAAG05889	Aag05889 Arabidops
50	141	7.0	247	3	AAAG38058	Aag38058 Arabidops
51	141	7.0	247	3	AAAG05888	Aag05888 Arabidops
52	141	7.0	263	3	AAAG38057	Aag38057 Arabidops
53	141	7.0	274	3	AAAG05887	Aag05887 Arabidops
54	141	7.0	463	7	ADD47587	Add47587 Rat prote
55	139.5	7.0	258	3	AAAB44500	Aab44500 Plant vir
56	138.5	6.9	827	5	ABP35602	Abp35602 Fungal ZB
57	135.5	6.8	177	4	ABG09342	Abg09342 Novel hum
58	135.5	6.8	1277	4	ABBE6688	Abbe6688 Drosophil
59	135	6.7	121	3	AAAG21511	Aag21511 Arabidops
60	134.5	6.7	172	4	ABG18832	Abg18832 Novel hum
61	133	6.6	144	4	AAAM42061	Aam42061 Human pol
62	132	6.6	369	7	ADD14178	Add14178 Human src
63	132	6.6	371	2	AAAW64535	Aaw64535 Human leu
64	132	6.6	371	5	ABP61444	Abp61444 Human NF-
65	132	6.6	371	6	ABR41017	Abra41017 Human MAP
66	131	6.5	101	3	AAAG21512	Aag21512 Arabidops
67	131	6.5	539	3	AAAB10224	Abi0224 Chicken n
68	130.5	6.5	237	6	ABER58292	Abx58292 BCUC371B
69	130.5	6.5	338	7	ADB75310	Adb75310 Prostate
70	130	6.5	277	3	AAAY44255	Aay44255 Monkey ap
71	130	6.5	308	3	AAAB44499	Aab44499 Plant vir
72	130	6.5	543	3	AAAB27551	Aab27551 Human tum
73	130	6.5	577	3	AAAB27550	Aab27550 Human tum
74	130	6.5	1646	3	AAAB27553	Aab27553 Human tum
75	130	6.5	1647	3	AAAB27549	Aab27549 Human tum
76	130	6.5	1649	3	AAAB27557	Aab27557 Human tum
77	130	6.5	1650	3	AAAB27555	Aab27555 Human tum
78	130	6.5	1678	3	AAAB27554	Aab27554 Human tum
79	130	6.5	1679	3	AAAB27552	Aab27552 Human tum
80	130	6.5	1681	3	AAAB27558	Aab27558 Human tum
81	130	6.5	1682	3	AAAB27556	Aab27556 Human tum
82	129.5	6.5	157	4	ABBE65626	Abbe65626 Drosophil
83	129.5	6.5	1419	5	AAO20572	Aao20572 Protein E
84	129.5	6.5	1419	6	ABBB99408	Abb99408 Amino aci
85	128.5	6.4	114	5	ABG77261	Abg77261 Selected
86	128.5	6.4	177	5	ABG11132	Abj11132 Yeast sel
87	128	6.4	612	4	AAAB95546	Aab95546 Human pro
88	128	6.4	1078	4	AAAM39300	Aam39300 Human pol
89	128	6.4	1085	7	ADD14093	Add14093 Human src
90	128	6.4	1088	4	AAAM80212	Aam80212 Human pro
91	128	6.4	1088	4	AAAM41086	Aam41086 Human pol
92	128	6.4	1089	4	ABG20503	Abg20503 Novel hum
93	128	6.4	1196	6	ABR39812	Abr39812 Human SCA
94	128	6.4	284	3	AAAY44254	Aay44254 Human apo
95	127.5	6.4	284	3	AAAB82706	Aab82706 Human apo
96	127.5	6.4	284	3	AAAB82706	Aab82706 Human apo
97	127.5	6.4	284	4	AAU29141	Aau29141 Human PRO
98	127.5	6.4	284	4	AAAB87556	Aab87556 Human PRO

99 127.5 6.4 284 4 AAB92943 Aab92943 Human pro
100 127.5 6.4 284 5 ABG95881 Abg95881 Human sec

ALIGNMENTS

```
RESULT 1
AAW19111
ID AAW19111 standard; protein; 380 AA.
AC AAW19111;
XX
DT 10-DEC-1997 (first entry)
XX
DE Human tumour susceptibility protein TSG101.
XX
KW Tumour susceptibility protein; TSG101; tumorigenesis; breast cancer;
KW diagnosis; therapy; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 11
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 38
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 38
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 44
FT /note= "potential N-glycosylation site"
FT Modified-site 86
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 89
FT /note= "potential protein kinase C phosphorylation site"
FT Domain 130..205
FT /note= "proline-rich domain"
FT Modified-site 150
FT /note= "potential N-glycosylation site"
FT Modified-site 210
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 215
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 225
FT /note= "potential protein kinase C phosphorylation site"
FT Domain 231..302
FT /note= "coiled-coil domain"
FT Modified-site 249
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 265
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 290
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 297
FT /note= "potential N-glycosylation site"
FT Modified-site 357
FT /note= "potential protein kinase C phosphorylation site"
XX
FN WO9718333-A1.
XX
XX 22-MAY-1997.
XX
XX 15-NOV-1996; 96WO-US018828.
XX
XX 16-NOV-1995; 95US-0006856P.
XX 16-JAN-1996; 96US-00585758.
XX 13-JUN-1996; 96US-00670274.
XX
XX (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
XX Li L, Cohen SN;
XX
XX WPI; 1997-289308/26.
```

DR N-PSDB; AAT69575.
XX Nucleic acid encoding tumour susceptibility protein TSG101 - useful to
FT develop products for cancer detection, prophylaxis, therapy and
PT predisposition determination.
XX
XX Example 1; Page 73-74; 89pp; English.
XX
CC This polypeptide sequence comprises a human tumour susceptibility
CC protein, designated TSG101, that can function as a tumour suppressor in
CC oncogenesis. Its amino acid sequence was deduced from a cDNA clone
CC (AAT69575) isolated from a human placenta cDNA library. It shows 94%
CC identity to human TSG101 (see AAW19110). Deletions in human TSG101 are
CC associated with the occurrence of cancers, e.g. breast cancer. Methods
CC are provided for identifying homologous or related proteins, for
CC producing compositions that modulate the expression or function of the
CC TSG101 protein, and for studying physiological pathways. Modulation of
CC gene activity in vivo is used for prophylactic or therapeutic purposes,
CC such as cancer therapy, identification of cell type based on expression,
CC etc
XX
SQ Sequence 380 AA;

```
Query Match 100.0%; Score 2002; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.4e-146;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSQYKYRDLTVRETNNVITLYKDLKPLVLDVSYVFDGSSRELMLNLTGTIPVYRGNTYNI 60
DB 1 MVSQYKYRDLTVRETNNVITLYKDLKPLVLDVSYVFDGSSRELMLNLTGTIPVYRGNTYNI 60
QY 61 PICLWLLDTVPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
DB 61 PICLWLLDTVPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
QY 121 MIVVFGDEPVPFGRPIASASYPYQATGPPNTSYMPGMPGGISYPYSGYPPNPSGYPGCPY 180
DB 121 MIVVFGDEPVPFGRPIASASYPYQATGPPNTSYMPGMPGGISYPYSGYPPNPSGYPGCPY 180
QY 181 PPGGYPATTSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKLRMKEEMDRAQ 240
DB 181 PPGGYPATTSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKLRMKEEMDRAQ 240
QY 241 AELNALKRTBEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKDKBELSSALEKMNQSE 300
DB 241 AELNALKRTBEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKDKBELSSALEKMNQSE 300
QY 301 NNDIDEVIPTAPLYKQILNLYAEENAIETIIFVLGEALRRGVIDLDFLKHVLLSRKQ 360
DB 301 NNDIDEVIPTAPLYKQILNLYAEENAIETIIFVLGEALRRGVIDLDFLKHVLLSRKQ 360
QY 361 FOLRALMQKARKTAGLSLDLY 380
DB 361 FOLRALMQKARKTAGLSLDLY 380

RESULT 2
AAW93424
ID AAW93424 standard; protein; 390 AA.
XX
XX AAW93424;
XX
XX 11-JUN-1999 (first entry)
XX
XX Human NHTS protein.
XX
XX Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
KW brain; breast; colon; heart; kidney; ovary; paragonia; pancreas;
KW prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
KW biliary cirrhosis; Crohn's disease; rheumatoid arthritis.
XX
XX Homo sapiens.
XX
```

PN US5892016-A.
XX
PD 06-APR-1999.
XX
XX 23-JAN-1997; 97US-00786999.
XX
XX 23-JAN-1997; 97US-00786999.
XX
XX (INCY-) INCYTE PHARM.
XX
XX Brie SL, Goli SK;
XX
XX WPI; 1999-253932/21.
XX
XX N-PSDB; AAX23168.
XX
XX Novel human tumor suppressor - useful for the diagnosis or treatment of
XX lymphoma, cancer, and autoimmune disease.
XX
XX Claim 1; Fig 1A-B; 25pp; English.
XX
XX This invention describes the isolation of a novel human tumor suppressor
XX (NHTS). The products of the invention may be used for the diagnosis or
XX treatment of conditions and diseases which are associated with expression
XX of NHTS e.g. lymphoma and cancers of the brain, breast, colon, heart,
XX kidney, ovary, paraganglia, pancreas, prostate, skin, stomach and thyroid
XX and autoimmune disease e.g. asthma, biliary cirrhosis, Crohn's disease,
XX diabetes, and rheumatoid arthritis
XX
SQ Sequence 390 AA;
Query Match 100.0%; Score 2002; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 7.7e-146; Indels 0; Gaps 0;
Matches 380; Conservative 0; Mismatches 0;
QY 1 MVSQKYRDLTVRETWNVITLYKDLKPVLDVSFVNDGSSRELNMNLTGTIPVYRGNTYNI 60
Db 11 MVSQKYRDLTVRETWNVITLYKDLKPVLDVSFVNDGSSRELNMNLTGTIPVYRGNTYNI 70
QY 61 PICMLLDTPYNPPICFVKPTSSMTIKTGKHDANGKIYLYLHFWKHPQSDLLGLIQV 120
Db 71 PICMLLDTPYNPPICFVKPTSSMTIKTGKHDANGKIYLYLHFWKHPQSDLLGLIQV 130
QY 121 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGISPYPSPGPPNPGSGYPCPY 180
Db 131 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGISPYPSPGPPNPGSGYPCPY 190
QY 181 PPGYPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
Db 191 PPGYPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250
QY 241 AELNALKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDDELSALEKVENQSE 300
Db 251 AELNALKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDDELSALEKVENQSE 310
QY 301 NNDIDEVITPAPLYKQILNLYAEENALIEDTIFYLGEALRGVIDLDVFLKHVRLLSRQ 360
Db 311 NNDIDEVITPAPLYKQILNLYAEENALIEDTIFYLGEALRGVIDLDVFLKHVRLLSRQ 370
QY 361 FQLRALMQARKTAGLSLDLY 380
Db 371 FQLRALMQARKTAGLSLDLY 390
RESULT 3
ABP52191
ID ABP52191 standard; protein; 390 AA.
XX
XX
AC ABP52191;
XX
XX
DT 15-OCT-2002 (first entry)
DE Human tumour susceptibility protein (TSG101).
XX

KW Human; tumour susceptibility protein; TSG101; proliferative disease; UD;
KW ubiquitination-regulating domain; MDM2; cytostatic.
XX
OS Homo sapiens.
XX
XX WO200257427-A2.
XX
XX 25-JUL-2002.
XX
XX 18-JAN-2002; 2002WO-US001947.
XX
XX 19-JAN-2001; 2001US-0262763P.
XX
XX (STRD) UNIV LELAND STANFORD JUNIOR.
XX
XX Cohen SN, Li L;
XX
XX WPI; 2002-590733/63.
XX
XX Novel antibody useful for regulating ubiquitination in cell by modulating
XX interaction of tumor susceptibility gene 101 with MDM2 proteins, binds
XX specifically to polypeptide having ubiquitination-regulating domain.
XX
XX Claim 9; Fig 6; 43pp; English.
XX
XX The present invention describes an antibody (I) that binds specifically
XX to a polypeptide (II) comprising an ubiquitination-regulating domain
XX (UD). Also described are: (1) a method (M1) for preparing (I); (2) a
XX method (M2) for treating a condition resulting from a change in a level
XX of MDM2 or tumor susceptibility gene (TSG)101 protein in cells of the
XX subject, involving administering to the subject an agent (A1) comprising
XX an UD or an agent (A2) that modulates the interaction of TSG101 protein
XX with MDM2, respectively; and (3) a method (M3) for treating a
XX proliferative disease in a subject involving monitoring the subject for
XX the level of p53 or TSG101, and treating the subject with A1 or A2 to
XX maintain the level of p53 or TSG101 within a target range, or
XX administering a therapeutically effective amount of A2 to the subject.
XX (I) has cytostatic activity. (M2) is useful for treating a condition
XX resulting from a change in a level of MDM2 or TSG101 protein in cells of
XX the subject. (M3) can be used for treating a proliferative disease in a
XX subject. (I) is useful for determining the interaction partner of
XX polypeptide comprising an UD, for measuring the level of polypeptide in a
XX sample, and in diagnostic methods. (I) is also useful for treating
XX proliferative disease. (I) regulates the ubiquitination in cells e.g. by
XX modulating the interaction between a TSG101 protein and an MDM2 protein.
XX The present sequence represents the human TSG101 protein which is used in
XX the exemplification of the present invention
XX
SQ Sequence 390 AA;
Query Match 100.0%; Score 2002; DB 5; Length 390;
Best Local Similarity 100.0%; Pred. No. 7.7e-146; Indels 0; Gaps 0;
Matches 380; Conservative 0; Mismatches 0;
QY 1 MVSQKYRDLTVRETWNVITLYKDLKPVLDVSFVNDGSSRELNMNLTGTIPVYRGNTYNI 60
Db 11 MVSQKYRDLTVRETWNVITLYKDLKPVLDVSFVNDGSSRELNMNLTGTIPVYRGNTYNI 70
QY 61 PICMLLDTPYNPPICFVKPTSSMTIKTGKHDANGKIYLYLHFWKHPQSDLLGLIQV 120
Db 71 PICMLLDTPYNPPICFVKPTSSMTIKTGKHDANGKIYLYLHFWKHPQSDLLGLIQV 130
QY 121 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGISPYPSPGPPNPGSGYPCPY 180
Db 131 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGISPYPSPGPPNPGSGYPCPY 190
QY 181 PPGYPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
Db 191 PPGYPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250
QY 241 AELNALKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDDELSALEKVENQSE 300
Db 251 AELNALKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDDELSALEKVENQSE 310

QY	301	NNDIDEVIITAPLYKQILNLYAEENAIETIYFLGEALRGVIDLDVFLKHVLLSRKQ	360
Db	311	NNDIDEVIITAPLYKQILNLYAEENAIETIYFLGEALRGVIDLDVFLKHVLLSRKQ	370
QY	361	FQLRALMOKARKTAGLSLDLY	380
Db	371	FQLRALMOKARKTAGLSLDLY	390
RESULT 4			
ABG73787	ID ABG73787 standard; protein; 390 AA.		
XX	AC	ABG73787;	
XX	10-APR-2003	(first entry)	
DT	DE	Human tumour suppressor NHTS.	
XX	Human; tumour suppressor; NHTS; cytostatic; antiasthmatic; hepatotropic; antiinflammatory; antidiabetic; antirheumatic; antiarthritic; cancer; autoimmune disease; asthma; biliary cirrhosis; Crohn's disease; diabetes; rheumatoid arthritis; drug screening; vaccine; gene therapy.		
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
FH	Region	256..277	
FT	/note= "Region containing leucine zipper motif"		
XX	US6472508-B1.		
XX	29-OCT-2002.		
XX	18-DEC-1998;	98US-00216387.	
XX	23-JAN-1997;	97US-00786999.	
XX	(INCY-) INCYTE GENOMICS INC.		
XX	La Brie S, Goli SK;		
XX	WPI; 2003-196850/19.		
DR	N-PSDB; ABQ77228.		
XX	New purified human tumor suppressor polypeptide, useful for diagnosing, preventing and treating cancer and autoimmune diseases (e.g. asthma or diabetes), and in drug screening assays.		
PS	Claim 1; Fig 1A-B; 26pp; English.		
XX	This invention describes a novel human tumour suppressor, NHTS, which has cytostatic, antiasthmatic, hepatotropic, antiinflammatory, antidiabetic, antirheumatic and antiarthritic activity. The NHTS polypeptide is useful in diagnosing, preventing and treating cancer and autoimmune diseases (e.g. asthma, biliary cirrhosis, Crohn's disease, diabetes or rheumatoid arthritis), in drug screening assays, in vaccines and for gene therapy. The nucleic acid encoding the above polypeptide may be used to detect and quantitate gene expression in biopsied tissues in which expression of the polypeptide may be correlated with disease, and thus, may be used to distinguish between absence, presence, and excess expression of the polypeptide, and to monitor the regulation of the polypeptide levels during therapeutic intervention. The polynucleotide may also be used to generate hybridisation probes for mapping genomic sequences, and in various molecular biology techniques. Antibodies generated from the products of the invention may be used as a targeting or delivery mechanism for bringing an agonist or other pharmaceutical agent to cells of patients having the above-mentioned diseases, and in diagnosing conditions or diseases characterised by the expression of the polypeptide, or in assays to monitor patients being treated with the above polypeptide, agonists or antagonists. This sequence represents the NHTS protein, isolated from Incyte clone 609476 derived from COLN0T01, a		
CC			

CC	colon cDNA library		
XX	Sequence 390 AA;		
SQ	Query Match	100.0%;	Score 2002; DB 6; Length 390;
	Best Local Similarity	100.0%;	Pred. No. 7.7e-146;
	Matches 380; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MVSKYKRDLTVRETNNVITLYKDLKPVLDVSYVNDGSSRELNNLTGTTIPVYRGNTYNI	60
Db	11	MVSKYKRDLTVRETNNVITLYKDLKPVLDVSYVNDGSSRELNNLTGTTIPVYRGNTYNI	70
QY	61	PICLWLLDTPYNPPICFVKPTSSMTIKTGKHVDANGKIYLYLHWHKHPQSDLLGLIQV	120
Db	71	PICLWLLDTPYNPPICFVKPTSSMTIKTGKHVDANGKIYLYLHWHKHPQSDLLGLIQV	130
QY	121	MIWVFGDEPPVFSRPISASYPYQATGPPNTSNMCPGGISYPSPGYPPNPSGYPGCPY	180
Db	131	MIWVFGDEPPVFSRPISASYPYQATGPPNTSNMCPGGISYPSPGYPPNPSGYPGCPY	190
QY	181	PRGGPYPATSSQYSPQPPVTTVGPGRDGTISEDTIRASLISAVSDKLRWKMKEMDRAQ	240
Db	191	PRGGPYPATSSQYSPQPPVTTVGPGRDGTISEDTIRASLISAVSDKLRWKMKEMDRAQ	250
QY	241	AELNALKRTTEEDLKKGKQKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKMNQSE	300
Db	251	AELNALKRTTEEDLKKGKQKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKMNQSE	310
QY	301	NNDIDEVIITAPLYKQILNLYAEENAIETIYFLGEALRGVIDLDVFLKHVLLSRKQ	360
Db	311	NNDIDEVIITAPLYKQILNLYAEENAIETIYFLGEALRGVIDLDVFLKHVLLSRKQ	370
QY	361	FQLRALMOKARKTAGLSLDLY	380
Db	371	FQLRALMOKARKTAGLSLDLY	390
RESULT 5			
ADCS5176	ID	ADCS5176 standard; protein; 390 AA.	
XX	AC	ADCS5176;	
XX	18-DEC-2003	(first entry)	
DT	DE	Novel human tumour suppressor, NHTS.	
XX	novel human tumour suppressor; NHTS; human; tumour; cancer; connective tissue disorder; rheumatoid arthritis.		
OS	Homo sapiens.		
XX	US2003099988-A1.		
XX	29-MAY-2003.		
XX	13-SEP-2002;	2002US-00243815.	
XX	23-JAN-1997;	97US-00786999.	
XX	18-DEC-1998;	98US-00216387.	
XX	(INCY-) INCYTE GENOMICS INC.		
XX	Brie SL, Goli SK;		
XX	WPI; 2003-687760/65.		
XX	N-PSDB; ADC35177, ADC35179.		
XX	Novel human tumor suppressor (NHTS) useful for preparing a composition for diagnosing, treating or preventing a disease or condition associated with expression of NHTS e.g. cancer.		
XX	Claim 1; SEQ ID NO 1; 29pp; English.		

XX The invention relates to an isolated polypeptide. The polypeptide is
 CC useful for preparing a composition for diagnosing or treating a disease
 CC or condition associated with decreased expression or overexpression of
 CC novel human tumour suppressor (NHTS) e.g. cancer or connective tissue
 CC disorders, such as rheumatoid arthritis. The present sequence represents
 CC the amino acid sequence of novel human tumour suppressor, NHTS.
 XX Sequence 390 AA;
 SQ
 Query Match 100.0%; Score 2002; DB 7; Length 390;
 Best Local Similarity 100.0%; Pred. No. 7.7e-146;
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVSQKYRDLTVRETQNVITLYKDLKPVLDSYVFNDSGSRRLMNLGTIPVYRGNTYNI 60
 Db |||||
 QY 61 PICLMLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGLIQV 120
 Db |||||
 QY 121 MIWFGDEPPVFSRPIASYPYQATGPNTSYMPGPGISYPYSGYPPNPSGYPGCPY 180
 Db |||||
 QY 131 MIWFGDEPPVFSRPIASYPYQATGPNTSYMPGPGISYPYSGYPPNPSGYPGCPY 190
 QY 181 PGGYPYATTSSQYPSQPPVTVTGPSRDGTISEDITRASLISAVSDKLRWKEEMDRAQ 240
 Db |||||
 QY 191 PGGYPYATTSSQYPSQPPVTVTGPSRDGTISEDITRASLISAVSDKLRWKEEMDRAQ 250
 QY 241 AELNALKRTEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQSE 300
 Db |||||
 QY 251 AELNALKRTEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQSE 310
 QY 301 NNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 360
 Db |||||
 QY 311 NNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 370
 Db |||||
 QY 361 FQLRALMOKARKTAGLSDLIY 380
 Db |||||
 QY 371 FQLRALMOKARKTAGLSDLIY 390
 Db |||||
 RESULT 6
 AA34885
 ID AA34885 standard; protein; 390 AA.
 AC AA34885;
 DT 28-MAY-2003 (first entry)
 DE Human Tsg101 protein.
 KW Human; twisted gastrulation; Tsg101; human immunodeficiency virus; HIV;
 KW gene therapy; peptide therapy.
 XX Homo sapiens.
 OS
 FN WO200294314-A1.
 XX
 PD 28-NOV-2002.
 XX
 PF 21-MAY-2002; 2002WO-US015965.
 XX
 PR 21-MAY-2001; 2001US-0292761P.
 XX
 PA (UYNV) UNIV NEW YORK STATE RES FOUND.
 XX
 PI Cohen SN, Carter C, Goff A, Ehrlich L;
 XX
 DR WPI; 2003-148440/14.
 XX
 PT Identifying twisted gastrulation 101 peptide, for treating human

PT immunodeficiency virus (HIV) infection, comprises comparing the level of
 PT HIV viral particles in a mammalian cell culture to that in a control
 XX culture.
 PS Claim 9; Col 45; 35pp; English.
 XX
 CC The invention relates to a method of identifying a mammalian twisted
 CC gastrulation (Tsg) 101 peptide. The method involves measuring the level
 CC of human immunodeficiency virus (HIV) viral particles released in a
 CC culture of mammalian cells having an expression construct comprising a
 CC portion of the coding sequence of a mammalian Tsg101 gene and comparing
 CC the level of HIV viral particles to that in a culture of control
 CC mammalian cells. The method is useful in identifying a peptide that is
 CC effective in reducing HIV particle production or which may be used in
 CC treating a patient infected with HIV or other retrovirus. The invention
 CC is useful in gene therapy and peptide therapy. The present sequence is
 CC human Tsg101 protein
 XX
 SQ Sequence 390 AA;

Query Match 99.8%; Score 1997; DB 6; Length 390;
 Best Local Similarity 99.7%; Pred. No. 1.9e-145;
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MVSQKYRDLTVRETQNVITLYKDLKPVLDSYVFNDSGSRRLMNLGTIPVYRGNTYNI 60
 Db |||||
 QY 11 MVSQKYRDLTVRETQNVITLYKDLKPVLDSYVFNDSGSRRLMNLGTIPVYRGNTYNI 70
 Db |||||
 QY 61 PICLMLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGLIQV 120
 Db |||||
 QY 71 PICLMLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGLIQV 130
 QY 121 MIWFGDEPPVFSRPIASYPYQATGPNTSYMPGPGISYPYSGYPPNPSGYPGCPY 180
 Db |||||
 QY 131 MIWFGDEPPVFSRPIASYPYQATGPNTSYMPGPGISYPYSGYPPNPSGYPGCPY 190
 QY 181 PGGYPYATTSSQYPSQPPVTVTGPSRDGTISEDITRASLISAVSDKLRWKEEMDRAQ 240
 Db |||||
 QY 191 PGGYPYATTSSQYPSQPPVTVTGPSRDGTISEDITRASLISAVSDKLRWKEEMDRAQ 250
 QY 241 AELNALKRTEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQSE 300
 Db |||||
 QY 251 AELNALKRTEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQSE 310
 QY 301 NNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 360
 Db |||||
 QY 311 NNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRQ 370
 QY 361 FQLRALMOKARKTAGLSDLIY 380
 Db |||||
 QY 371 FQLRALMOKARKTAGLSDLIY 390
 Db |||||

RESULT 7
 AA19110
 ID AA19110 standard; protein; 381 AA.
 XX
 AC AA19110;
 XX
 DT 10-DEC-1997 (first entry)
 XX
 DE Mouse tumour susceptibility protein TSG101.
 XX
 KW Tumour susceptibility protein; TSG101; tumorigenesis; breast cancer;
 KW diagnosis; therapy; mouse.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 11
 FT /note= "potential protein kinase C phosphorylation site"
 FT Region 37..46
 FT /note= "residues 37-46 resemble a helix-turn-helix

FT Modified-site 38 signature domain"
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 38
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 44
FT /note= "potential N-glycosylation site"
FT Modified-site 55
FT /note= "potential N-myristoylation site"
FT 73. .83
FT /note= "residues 73-83 resemble a fungal Zn-cys bi-
FT nuclear cluster signature "
FT 85
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 88
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 150
FT /note= "potential N-glycosylation site"
FT Modified-site 156
FT /note= "potential N-myristoylation site"
FT Modified-site 210
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 215
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 225
FT /note= "potential protein kinase C phosphorylation site"
FT Modified-site 249
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 265
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 290
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 297
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 357
FT /note= "potential protein kinase C phosphorylation site"
XX
PN W09718333-A1.
XX
XX 22-MAY-1997.
XX
XX 15-NOV-1996; 96WO-US018828.
XX
XX 16-NOV-1995; 95US-0006856P.
XX 16-JAN-1996; 96US-00585758.
XX 13-JUN-1996; 96US-00670274.
XX
XX (STRD) UNIV LELAND STANFORD JUNIOR.
XX
XX Li L, Cohen SN;
XX
XX WPI; 1997-289308/26.
XX N-PSDB; AAT69574.
XX
XX Nucleic acid encoding tumour susceptibility protein TSG101 - useful to
XX develop products for cancer detection, prophylaxis, therapy and
XX predisposition determination.
XX
XX Example 1; Page 67-70; 89pp; English.
XX
XX This polypeptide sequence comprises a mouse tumour susceptibility
XX protein, designated TSG101, that can function as a tumour suppressor in
XX oncogenesis. Its amino acid sequence was deduced from a cDNA clone
XX (AAT69574) isolated using a novel random homoygous knockout approach. It
XX shows 94% identity to human TSG101 (see AAM19111). Deletions in human
XX TSG101 are associated with the occurrence of cancers, e.g. breast cancer.
XX Methods are provided for identifying homologous or related proteins, for
XX producing compositions that modulate the expression or function of the
XX TSG101 protein, and for studying physiological pathways. Modulation of
XX gene activity in vivo is used for prophylactic or therapeutic purposes.
XX such as cancer therapy, identification of cell type based on expression,
XX etc
XX

SQ Sequence 381 AA;
Query Match 94.9%; Score 1900.5; DB 2; Length 381;
Best Local Similarity 94.5%; Pred. No. 5e-138;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MVSQYKRDLTRETQVNVITLYKDLKPVLDSYVFNDSGSSRELNMNLTGTTPVYRGNTYNI 60
DB 1 MMSKYKRDLTVRQTVNVIAAMYKDLAPVLDSYVFNDSGSSRELNVLTGTTPVYRGNTYNI 60
QY 61 PICLWLLDTPYNNPPICFVKPSSMTIKTKGKVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
DB 61 PICLWLLDTPYNNPPICFVKPSSMTIKTKGKVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
QY 121 MIVVFGDEBPVRSRP-ISAISYPPYQATGPNTSYMFGMPGGISPYPSGYPPNPSGYPGCP 179
DB 121 MIVIFGEEBPVRSRPVTSASYPPTATGPNTSYMFGMPGGISPYPSGYPPNPSGYPGCP 180
QY 180 YPPGPGYPATTSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWKEEMDR 239
DB 181 YPPAGYPATTSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWKEEMDR 240
QY 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQS 299
DB 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQS 300
QY 300 ENNDIDBVIPTAPLYKQILNLYAEENAIEDTTFYLGELRRGVVIDLDFLKHVRLLSRK 359
DB 301 ENNDIDBVIPTAPLYKQILNLYAEENAIEDTTFYLGELRRGVVIDLDFLKHVRLLSRK 360
QY 360 QFQLRALMQKARKTAGLSDIY 380
DB 361 QFQLRALMQKARKTAGLSDIY 381
RESULT 8
AAW93425
ID AAW93425 standard; protein; 381 AA.
XX
XX AC AAW93425;
XX
XX 11-JUN-1999 (first entry)
XX
XX Mouse tsgl01 protein.
XX
XX Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
XX brain; breast; colon; heart; kidney; ovary; paraganglia; pancreas;
XX prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
XX biliary cirrhosis; Crohn's disease; rheumatoid arthritis; mouse; tsgl01.
XX
XX Mus sp.
XX
XX US5892016-A.
XX
XX 06-APR-1999.
XX
XX 23-JAN-1997; 97US-00786999.
XX
XX 23-JAN-1997; 97US-00786999.
XX
XX (INCY-) INCYTE PHARM.
XX
XX Brie SL, Goli SK;
XX
XX WPI; 1999-253932/21.
XX
XX Novel human tumor suppressor - useful for the diagnosis or treatment of
XX lymphoma, cancer, and autoimmune disease.
XX
XX Disclosure; Fig 2; 25pp; English.
XX
XX This invention describes the isolation of a novel human tumour suppressor
XX (NHTS). The products of the invention may be used for the diagnosis or
XX

CC treatment of conditions and diseases which are associated with expression
CC of NHTS e.g. lymphoma and cancers of the brain, breast, colon, heart,
CC kidney, ovary, paraganglia, pancreas, prostate, skin, stomach and thyroid
CC and autoimmune disease e.g. asthma, biliary cirrhosis, Crohn's disease,
CC diabetes, and rheumatoid arthritis
XX
SQ Sequence 381 AA;

Query Match 94.9%; Score 1900.5; DB 2; Length 381;
Best Local Similarity 94.5%; Pred. No. 5e-138;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MVSKEYKRDLTVRETVNVTLLKDKLPVLDVYVNDGSSRELNMNLTGTPVYRGNTYNI 60
Db 1 MMSKYKRDLTVRQTVNVIAMTKDKLPVLDVYVNDGSSRELNMNLTGTPVYRGNTYNI 60
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDKHQPQSDLLGLIQV 120
Db 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDKHQPQSDLLGLIQV 120
QY 121 MIVFGDEPPVFSRP- ISASYPPYQATGPPNTSYMPGPGGSIYPSGYPNPGSGYPCP 179
Db 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGPGGSIYPSGYPNPGSGYPCP 180
QY 180 YPPGGYPATTSQVPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWKEEMDRA 239
Db 181 YPPAGYPATTSQVPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWKEEMDRA 240
QY 240 QAEINLAKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDEELSSALEKWNQS 299
Db 241 QAEINLAKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDEELSSALEKWNQS 300
QY 300 ENNDIDEVILITAPLYKQILNLYAEENAIETIFYLGEALRGVTDLDVFLKHVRLLSRK 359
Db 301 ENNDIDEVILITAPLYKQILNLYAEENAIETIFYLGEALRGVTDLDVFLKHVRLLSRK 360
QY 360 QFQLRALMOKARKTAGLSLDLY 380
Db 361 QFQLRALMOKARKTAGLSLDLY 381

RESULT 9

ABG73788
ID ABG73788 standard; protein; 381 AA.
XX
AC ABG73788;

XX 10-APR-2003 (first entry)

XX Murine tsig101 protein.

KW Murine; tumour suppressor; NHTS; cytostatic; antiasthmatic; hepatotropic;
KW antiinflammatory; antidiabetic; antirheumatic; antiarthritic; cancer;
KW autoimmune disease; asthma; biliary cirrhosis; Crohn's disease; diabetes;
KW rheumatoid arthritis; drug screening; vaccine; gene therapy; tsig101.

OS Mus musculus.

XX US6472508-B1.

XX 29-OCT-2002.

XX 18-DEC-1998; 98US-00216387.

XX 23-JAN-1997; 97US-00786999.

XX (INCY-) INCYTE GENOMICS INC.

XX La Brie S, Goli SK;

XX WPI; 2003-196850/19.

XX New purified human tumor suppressor polypeptide, useful for diagnosing,

PT preventing and treating cancer and autoimmune diseases (e.g. asthma or
PT diabetes), and in drug screening assays.

PS Disclosure; Fig 2; 26pp; English.

XX This invention describes a novel human tumor suppressor, NHTS, which has
CC cytostatic, antiasthmatic, hepatotropic, antiinflammatory, antidiabetic,
CC antirheumatic and antiarthritic activity. The NHTS polypeptide is useful
CC in diagnosing, preventing and treating cancer and autoimmune diseases
CC (e.g. asthma, biliary cirrhosis, Crohn's disease, diabetes or rheumatoid
CC arthritis), in drug screening assays, in vaccines and for gene therapy.
CC The nucleic acid encoding the above polypeptide may be used to detect and
CC quantitate gene expression in biopsied tissues in which expression of the
CC polypeptide may be correlated with disease, and thus, may be used to
CC distinguish between absence, presence, and excess expression of the
CC polypeptide, and to monitor the regulation of the polypeptide levels
CC during therapeutic intervention. The polynucleotide may also be used to
CC generate hybridisation probes for mapping genomic sequences, and in
CC various molecular biology techniques. Antibodies generated from the
CC products of the invention may be used as a targeting or delivery
CC mechanism for bringing an agonist or other pharmaceutical agent to cells
CC of patients having the above-mentioned diseases, and in diagnosing
CC conditions or diseases characterised by the expression of the
CC polypeptide, or in assays to monitor patients being treated with the
CC above polypeptide, agonists or antagonists. This sequence represents the
CC murine tsig101 protein, a homologue of the human NHTS protein, isolated
CC from Incyte clone 609476 derived from COLNNOT01, a colon cDNA library
XX

SQ Sequence 381 AA;

Query Match 94.9%; Score 1900.5; DB 6; Length 381;

Best Local Similarity 94.5%; Pred. No. 5e-138;

Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVSKEYKRDLTVRETVNVTLLKDKLPVLDVYVNDGSSRELNMNLTGTPVYRGNTYNI 60

Db 1 MMSKYKRDLTVRQTVNVIAMTKDKLPVLDVYVNDGSSRELNMNLTGTPVYRGNTYNI 60

QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDKHQPQSDLLGLIQV 120

Db 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLYLHDKHQPQSDLLGLIQV 120

QY 121 MIVFGDEPPVFSRP- ISASYPPYQATGPPNTSYMPGPGGSIYPSGYPNPGSGYPCP 179

Db 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGPGGSIYPSGYPNPGSGYPCP 180

QY 180 YPPGGYPATTSQVPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWKEEMDRA 239

Db 181 YPPAGYPATTSQVPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWKEEMDRA 240

QY 240 QAEINLAKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDEELSSALEKWNQS 299

Db 241 QAEINLAKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKDEELSSALEKWNQS 300

QY 300 ENNDIDEVILITAPLYKQILNLYAEENAIETIFYLGEALRGVTDLDVFLKHVRLLSRK 359

Db 301 ENNDIDEVILITAPLYKQILNLYAEENAIETIFYLGEALRGVTDLDVFLKHVRLLSRK 360

QY 360 QFQLRALMOKARKTAGLSLDLY 380

Db 361 QFQLRALMOKARKTAGLSLDLY 381

RESULT 10

ADG35178
ID ADG35178 standard; protein; 381 AA.

XX ADG35178;

XX 18-DEC-2003 (first entry)

XX Human tsig101.

KW novel human tumour suppressor; NHTS; human; tumour; cancer;
KW connective tissue disorder; rheumatoid arthritis; tsig101.
XX Homo sapiens.
XX US2003099988-A1.
XX 29-MAY-2003.
XX 13-SEP-2002; 2002US-00243815.
XX 23-JAN-1997; 97US-00786999.
PR 18-DEC-1998; 98US-00216387.
XX (INCY-) INCYTE GENOMICS INC.
XX PA Brie SL, Goli SK;
XX WIPI; 2003-687760/65.
XX Novel human tumor suppressor (NHTS) useful for preparing a composition
PT for diagnosing, treating or preventing a disease or condition associated
PT with expression of NHTS e.g. cancer.
XX Disclosure; SEQ ID NO 3; 29pp; English.
XX The invention relates to an isolated polypeptide. The polypeptide is
CC useful for preparing a composition for diagnosing or treating a disease
CC or condition associated with decreased expression or overexpression of
CC novel human tumour suppressor (NHTS) e.g. cancer or connective tissue
CC disorders, such as rheumatoid arthritis. The present sequence represents
CC the amino acid sequence of human tsig101.
XX
SQ Sequence 381 AA;
Query Match 94.9%; Score 1900.5; DB 7; Length 381;
Best Local Similarity 94.5%; Pred. No. 5e-138;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
Qy 1 MVSXKYRDLTVRTVNVVITLYKDKLPVDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
Db 1 MMSKYKYRDLTVRQTUNVVIAMKYDKLPVDSYVFNDSGSSRELNVLTGTIPVYRGNTYNI 60
Qy 61 PICLWLDTPYNNPICFVKPTSSMTIKTKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 61 PICLWLDTPYNNPICFVKPTSSMTIKTKHVDANGKIYLPYLHDKHPRSELELLIQI 120
Qy 121 MIVVFGDEPPVFSRP-ISA5YPPYQATGPPNTSYMPGPGISYPSPGYPNPSPGPGCP 179
Db 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGPGISAYSYPSPGYPNPSPGPGCP 180
Qy 180 YPPGPGYPATSSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSKLRWRMKEEMDRA 239
Db 181 YPPAGPYPATSSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSKLRWRMKEEMDGA 240
Qy 240 QAEALNALKRTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQS 299
Db 241 QAEALNALKRTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQS 300
Qy 300 ENNDIDEVITPTAPLYKQILNLYAEENAIETIFVLGEALRRGVLDLDFLKHVRLLSRK 359
Db 301 ENNDIDEVITPTAPLYKQILNLYAEENAIETIFVLGEALRRGVLDLDFLKHVRLLSRK 360
Qy 360 QFQRLALMQARKTAGLSLDY 380
Db 361 QFQRLALMQARKTAGLSLDY 381
RESULT 11
AAE34884
ID AAE34884 standard; protein; 391 AA.
XX
AC AAE34884;
XX 28-MAY-2003 (first entry)
XX Mouse Tsgl01 protein.
XX Mouse; twisted gastrulation; Tsgl01; human immunodeficiency virus; HIV;
KW gene therapy; peptide therapy.
XX Mus sp.
XX WO200294314-A1.
XX 28-NOV-2002.
XX 21-MAY-2002; 2002WO-US015965.
XX 21-MAY-2001; 2001US-0292761P.
XX (UYNY) UNIV NEW YORK STATE RES FOUND.
XX Cohen SN, Carter C, Goff A, Ehrlich L;
XX WIPI; 2003-148440/14.
XX Identifying twisted gastrulation 101 peptide, for treating human
PT immunodeficiency virus (HIV) infection, comprises comparing the level of
PT HIV viral particles in a mammalian cell culture to that in a control
PT culture.
XX Claim 9; Col 44; 35pp; English.
XX The invention relates to a method of identifying a mammalian twisted
CC gastrulation (Tsg) 101 peptide. The method involves measuring the level
CC of human immunodeficiency virus (HIV) viral particles released in a
CC culture of mammalian cells having an expression construct comprising a
CC portion of the coding sequence of a mammalian Tsg101 gene and comparing
CC the level of HIV viral particles to that in a culture of control
CC mammalian cells. The method is useful in identifying a peptide that is
CC effective in reducing HIV particle production or which may be used in
CC treating a patient infected with HIV or other retrovirus. The invention
CC is useful in gene therapy and peptide therapy. The present sequence is
CC mouse Tsg101 protein
XX
SQ Sequence 391 AA;
Query Match 94.9%; Score 1900.5; DB 6; Length 391;
Best Local Similarity 94.5%; Pred. No. 5.2e-138;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
Qy 1 MVSXKYRDLTVRTVNVVITLYKDKLPVDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
Db 11 MMSKYKYRDLTVRQTUNVVIAMKYDKLPVDSYVFNDSGSSRELNVLTGTIPVYRGNTYNI 70
Qy 61 PICLWLDTPYNNPICFVKPTSSMTIKTKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 71 PICLWLDTPYNNPICFVKPTSSMTIKTKHVDANGKIYLPYLHDKHPRSELELLIQI 130
Qy 121 MIVVFGDEPPVFSRP-ISA5YPPYQATGPPNTSYMPGPGISYPSPGYPNPSPGPGCP 179
Db 131 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGPGISAYSYPSPGYPNPSPGPGCP 190
Qy 180 YPPGPGYPATSSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSKLRWRMKEEMDRA 239
Db 191 YPPAGPYPATSSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSKLRWRMKEEMDGA 250
Qy 240 QAEALNALKRTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQS 299
Db 251 QAEALNALKRTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKKDEELSSALEKMNQS 310
Qy 300 ENNDIDEVITPTAPLYKQILNLYAEENAIETIFVLGEALRRGVLDLDFLKHVRLLSRK 359
Db 311 ENNDIDEVITPTAPLYKQILNLYAEENAIETIFVLGEALRRGVLDLDFLKHVRLLSRK 370

QY 360 QFQRLMOKARKTAGLSLDLY 380
 Db 371 QFQRLMOKARKTAGLSLDLY 391

RESULT 12
 ADB85228
 ID ADB85228 standard; protein; 391 AA.
 XX
 AC ADB85228;
 XX
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Mouse tumour susceptibility tsg101 SEQ ID NO:109.
 XX
 KW mouse; streptozocin; kinase; phosphatase; ion channel protein; receptor;
 KW transporter; G-protein coupled receptor; GPCR; DNA-binding proteins;
 KW protease; enzyme; analgesic; gene therapy; pain; diabetes.
 XX
 OS Mus musculus.
 XX
 XX
 PN EP1284297-A2.
 XX
 XX
 PD 19-FEB-2003.
 XX
 PF 26-JUL-2002; 2002EP-00255228.
 XX
 PR 27-JUL-2001; 2001GB-00018354.
 XX
 PR 07-FEB-2002; 2002GB-00002880.
 XX
 PA (WARN) WARNER LAMBERT CO.
 XX
 PI Brooksbank RA, Dixon AK, Lee K, Pincock RD;
 XX
 DR WPI; 2003-364994/35.
 DR N-PSDB; ADB85229.
 XX

Use of gene sequence that is down-regulated in response to streptozocin-induced diabetes, vector, host cell, animal, polypeptide and antibody, in screening of compounds for treating or diagnosing pain.

Disclosure; Page 175-176; 256pp; English.

The invention relates to a novel isolated gene sequence that is down-regulated in the spinal cord in response to streptozocin-induced diabetes, or comprising, hybridising or having at least 80% sequence identity to a sequence whose expression products are kinases, phosphatases, ion channel proteins, receptors, transporters, G-protein coupled receptor proteins, DNA-binding proteins, proteases or enzymes, given in the specification. A gene of the invention has analgesic activity, and may have a use in gene therapy. The gene sequences, vector, host cell, animal, polypeptide and antibody are useful for screening of compounds for diagnosing or treating pain. The kits are useful for simultaneous, separate or sequential detecting and/or quantifying down-regulation of a gene sequence in the spinal cord of a mammal in response to streptozocin-induced diabetes. The compound or pharmaceutical composition is useful as a medicament for treating or diagnosing pain. The present sequence represents a protein encoded by a gene of the invention.

Sequence 391 AA;
 SQ

Query Match 94.9%; Score 1900.5; DB 7; Length 391;
 Best Local Similarity 94.5%; Pred. NO. 5.2e-138;
 Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVSXKYRDLAVRETQNVITLYKDLKPVLDGVFNDGSSRELNLGTGTPVYRGNTVNI 60
 Db 11 MMSXKYRDLTVRTQNVIVAMKDKLPVLDGVFNDGSSRELNLGTGTPVYRGNTVNI 70
 QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHVKHQPQSLGLITQV 120
 Db 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHVKHQPQSLGLITQV 130

QY 121 MIVVFGDEPVPFGRP-ISASYPYPYQATGPNTSYMPGMPGSIPIPSGYPPNPSGYPGCP 179
 Db 131 MIVIFGEEPPVFSRPTVSASYPYPTATGPNTSYMPGMPGSIPIPSGYPPNPSGYPGCP 190
 QY 180 YPPGGYPATNTSSQYPSQPPVTTVGPSPRDTISSETIRASLISAVSDKLRWRKMEEMDRA 239
 Db 191 YPPAGYPATNTSSQYPSQPPVTTVGPSPRDTISSETIRASLISAVSDKLRWRKMEEMDGA 250
 QY 240 QAEINLAKKTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKMEENQS 299
 Db 251 QAEINLAKKTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKMEENQS 310
 QY 300 ENNDIDEVIPTAPLYKQILNLYAEENALIEDTIFYLGEALRRGVDDLDVFLKHVRLLSRK 359
 Db 311 ENNDIDEVIPTAPLYKQILNLYAEENALIEDTIFYLGEALRRGVDDLDVFLKHVRLLSRK 370
 QY 360 QFQRLMOKARKTAGLSLDLY 380
 Db 371 QFQRLMOKARKTAGLSLDLY 391

RESULT 13
 ABP41729
 ID ABP41729 standard; protein; 307 AA.
 XX
 AC ABP41729;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HSPSE88, SEQ ID NO:2861.
 XX
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive; chromosome 11p15.

Homo sapiens.
 WO200200677-A1.
 03-JAN-2002.
 07-JUN-2001; 2001WO-US018569.
 07-JUN-2000; 2000US-0209467P.
 (HUMA-) HUMAN GENOME SCI INC.
 Birse CE, Rosen CA;
 WPI; 2002-147878/19.
 N-PSDB; ABQ54806.
 Isolated nucleic acid molecules encoding novel ovarian polypeptides, useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian cancer), immune disorders, cardiovascular disorders and neurological diseases.

Claim 11; SEQ ID NO 2861; 2922pp; English.
 The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use

CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer,
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 307 AA;

Query Match 70.1%; Score 1404; DB 5; Length 307;
 Best Local Similarity 100.0%; Pred. No. 7.1e-100;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 110 PQSDLLGLIQVMIVFGDEPPVFSRPIASVPPYQATGPNTSYMPGMPGGISYPGYP 169
 Db 37 PQSDLLGLIQVMIVFGDEPPVFSRPIASVPPYQATGPNTSYMPGMPGGISYPGYP 96
 QY 170 PNPSPGPGCPYPGPGYPATTSQYSPQPPVTTVGPSSRDGTISEDTIRASLISAVSKLR 229
 Db 97 PNPSPGPGCPYPGPGYPATTSQYSPQPPVTTVGPSSRDGTISEDTIRASLISAVSKLR 156
 QY 230 WRMKEMDRAQAEALNALKRTBEDLKGHQKLEEMVTRLDQEAEDVKNIELKKDEELS 289
 Db 157 WRMKEMDRAQAEALNALKRTBEDLKGHQKLEEMVTRLDQEAEDVKNIELKKDEELS 216
 QY 290 SALEKEMENQSENNIDIVIIPTAPLYKQILNLYAENAIETIFVLGEALRRGVLDLDFV 349
 Db 217 SALEKEMENQSENNIDIVIIPTAPLYKQILNLYAENAIETIFVLGEALRRGVLDLDFV 276
 QY 350 LKHVRLLSRKQFQALRMQKARKTAGLSLDLY 380
 Db 277 LKHVRLLSRKQFQALRMQKARKTAGLSLDLY 307

RESULT 14
 AAE09328
 ID AAE09328 standard; protein; 237 AA.

XX AC AAE09328;
 XX DT 19-NOV-2001 (first entry)
 XX DE Human intracellular regulatory molecule, tsig101.

XX KW Human; intracellular regulator; cell division; proliferation; therapy;
 XX KW cancer; infection; wound; developmental abnormality; metabolic problem;
 XX KW cytostatic; antibacterial; vulnery; transcription factor; tsig101.

XX OS Homo sapiens.
 XX XX US6274312-B1.
 XX PD 14-AUG-2001.

XX PF 10-DEC-1997; 97US-0099977A.

PR 11-DEC-1996; 96US-0032818P.
 XX (SCHE) SCHERING CORP.
 XX Gish KC, Seghezzi W, Shanahan F, Lees EM, Mcclanahan TK;
 XX WPI; 2001-535086/59.
 DR N-PSDB; AAD16104.
 XX New genes encoding intracellular regulatory molecules, useful for
 PT regulating cell division and proliferation (e.g. tumor cells),
 PT particularly for treating cancer, infections, wounds, or developmental or
 PT metabolic abnormalities.
 XX Example 2; Col 29-32; 62pp; English.
 XX The present sequence is a human intracellular regulatory molecule,
 CC tsig101. The tsig101 is a transcription factor. The polynucleotides
 CC encoding intracellular regulatory molecules are useful for regulating
 CC cell division and proliferation of various cell types, including tumour
 CC cells. Specifically, they are also useful for treating cancer,
 CC infections, wounds, developmental abnormalities or metabolic problems
 XX SQ Sequence 237 AA;

Query Match 55.8%; Score 1117; DB 4; Length 237;
 Best Local Similarity 98.0%; Pred. No. 6.7e-78;
 Matches 201; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MVSKYKYRDLTVRETNVITLYKDLKPVLDSEYFNDGSSRELMLNLTGTIPVYRGNTYNI 60
 Db 11 MVSKYKYRDLTVRETNVITLYKDLKPVLDSEYFNDGSSRELMLNLTGTIPVYRGNTYNI 70
 QY 61 PICLWLDDTTPYAPPICFVKTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
 Db 71 PICLWLDDTTPYAPPICFVKTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130
 QY 121 MIVVFGDEPPVFSRPIASVPPYQATGPNTSYMPGMPGGISYPGYPNPSPGPGCPY 180
 Db 131 MIVVFGDEPPVFSRPIASVPPYQATGPNTSYMPGMPGGISYPGYPNPSPGPGCPY 190
 QY 181 PPGGYPATTSQYSPQPPVTTVGP 205
 Db 191 PPGGYPATTSQYSPQPPVTTVGP 215

RESULT 15
 ABB64607
 ID ABB64607 standard; protein; 331 AA.

XX AC ABB64607;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 20613.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.

XX OS Drosophila melanogaster.
 XX XX WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

```
XX WPI: 2001-656860/75.
DR N-PSDB; ABL08710.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 20613; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 331 AA;
XX
Query Match 40.4%; Score 809.5; DB 4; Length 331;
Best Local Similarity 50.0%; Pred. No. 5.3e-54;
Matches 168; Conservative 51; Mismatches 86; Indels 31; Gaps 7;
QY 67 LDTYPNPICFVKPTSSMTIKTGHVDANGKIYLPYLHEWKHPQSDLLGLIQWMIIVFG 126
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MDTHPQNAPMCFVKPTTMTQIKVSMYVDHNGKVLPLYLHDWPHSSDLLSLIQWMIIVTG 60
QY 127 DPPVFSRPISASYPYQATGPNTSYM--PGMPGGS---PYP-----SGYPPNPSG 174
Db :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 DHPVTSRKEQIAAPY-----FTSYMPQAPGGSNLSFLPYPTAGGAGSNFPFPYPTG 115
QY 175 YPGCFYPP-----GGPYPA-----TTSSQVPSQPPVTVTPGSRDGTISDITRASLI 221
Db :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 SNVGYPPTPAGPAGSGVPAYPNFIQPTAGGYPAAGNPNPNPSTGTITTEHFKASII 175
QY 222 SAVSKLKWKMEKMDRAQAEINALKRTTEDLKKGHKLKEEMVTRLDQAEVVDKNIELL 281
Db :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 SAIDDKLRRVQEKVNOYQAEITLNRTKQELLEGSAKIDAIERLERHIDMQKNISIL 235
QY 282 KKDEELSSALEKMNSENNDIDEVITPTAPLYKQILNLYAEENAIETPTYLGEALRR 341
Db :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 KDKEELKALEDLSEAINP-DEAVTTTAPLYQLLNAYADEAATAEDAIYYLGEGLRG 294
QY 342 GVIDLDVFLKHVRLSRKQFQLRALMQKARKTAGLS 377
Db :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 GVIDLETFLKHVRLSRKQFILRALMQKCRQKAGLA 330
RESULT 16
AAG06370
ID AAG06370 standard; protein; 398 AA.
AC AAG06370;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 3119.
XX
XX protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
99US-0121825P.
99US-0123180P.
99US-0123548P.
99US-0125788P.
99US-0126264P.
99US-0126785P.
99US-0127462P.
99US-0128234P.
99US-0128714P.
99US-0129845P.
99US-0130077P.
99US-0130449P.
99US-0130510P.
99US-0130891P.
99US-0131449P.
99US-0132048P.
99US-0132407P.
99US-0132484P.
99US-0132485P.
99US-0132486P.
99US-0132487P.
99US-0132863P.
99US-0134256P.
99US-0134218P.
99US-0134219P.
99US-0134221P.
99US-0134370P.
99US-0134768P.
99US-0134941P.
99US-0135124P.
99US-0135353P.
99US-0135629P.
99US-0136021P.
99US-0136392P.
99US-0136782P.
99US-0137222P.
99US-0137528P.
99US-0137502P.
99US-0137724P.
99US-0138094P.
99US-0138540P.
99US-0138847P.
99US-0139119P.
99US-0139452P.
99US-0139453P.
99US-0139454P.
99US-0139455P.
99US-0139456P.
99US-0139457P.
99US-0139458P.
99US-0139459P.
99US-0139460P.
99US-0139461P.
99US-0139462P.
99US-0139463P.
99US-0139750P.
99US-0139763P.
99US-0139817P.
99US-0139899P.
99US-0140353P.
99US-0140354P.
99US-0140695P.
99US-0140823P.
99US-0140991P.
99US-0141287P.
99US-0141842P.
99US-0142154P.
99US-0142055P.
99US-0142390P.
99US-0142803P.
99US-0142920P.
```


PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139817P.
PR	21-JUN-1999;	99US-0139899P.
PR	22-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	22-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161952P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 21.2%; Score 424; DB 3; Length 340;		
Best Local Similarity 32.2%; Pred. No. 2.9e-24;		
Matches 111; Conservative 63; Mismatches 139; Indels 32; Gaps 10;		
Qy	35	NDGSRRLMLTGTIPVPGNTYNIPICLWLLTPYNNPICFVKPTSSMTIK-TGKHV 93
Db	3	NDGSRVLLQADGTIPMPFHGVTNIPVILLESYPRHPPCVVYVNTADMIIKRPHAV 62
Qy	94	DANGKIYLPYLHEWKHPQSDLLGLIQVMIVVGDEPPVFSRPSIASYPPYQATGPPNT-- 151

PR	07-JUL-2000;	2000US-0216647P.	PR	20-OCT-2000;	2000US-0241808P.
PR	07-JUL-2000;	2000US-0216880P.	PR	20-OCT-2000;	2000US-0241809P.
PR	11-JUL-2000;	2000US-0217487P.	PR	20-OCT-2000;	2000US-0241826P.
PR	11-JUL-2000;	2000US-0217496P.	PR	01-NOV-2000;	2000US-0244617P.
PR	14-JUL-2000;	2000US-0218290P.	PR	08-NOV-2000;	2000US-0246474P.
PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246611P.
PR	18-AUG-2000;	2000US-0226279P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226686P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249277P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232081P.	PR	17-NOV-2000;	2000US-0249300P.
PR	12-SEP-2000;	2000US-0231968P.	PR	01-DEC-2000;	2000US-0250160P.
PR	14-SEP-2000;	2000US-0232397P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232400P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232401P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251869P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0234274P.	PR	08-DEC-2000;	2000US-0251990P.
PR	25-SEP-2000;	2000US-0234977P.	PR	11-DEC-2000;	2000US-0254097P.
PR	25-SEP-2000;	2000US-0235484P.	PR	05-JAN-2001;	2001US-0259678P.
PR	27-SEP-2000;	2000US-0235834P.	XX	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235836P.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327P.	XX	WPI; 2001-465566/50.	
PR	29-SEP-2000;	2000US-0236367P.	XX	N-PSDB; AAS40986.	
PR	29-SEP-2000;	2000US-0236368P.	XX	Novel polypeptides and polynucleotides useful for diagnosing, preventing,	
PR	29-SEP-2000;	2000US-0236369P.	PT	treating neural, immune system, muscular, reproductive, pulmonary,	
PR	29-SEP-2000;	2000US-0236370P.	PT	cardiovascular, renal, proliferative disorders and cancerous diseases.	
PR	02-OCT-2000;	2000US-0236802P.	XX	Claim 11; SEQ ID NO 1112; 1180pp; English.	
PR	02-OCT-2000;	2000US-0237037P.	PS	The present invention relates to the isolation of novel human enzyme	
PR	02-OCT-2000;	2000US-0237038P.	XX	polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences	
PR	02-OCT-2000;	2000US-0237039P.	CC	encoding them. The enzyme polypeptides of the invention may comprise the	
PR	13-OCT-2000;	2000US-0239935P.	CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,	
PR	20-OCT-2000;	2000US-0240960P.	CC	isomerases or ligases. The sequences of the invention are useful in the	
PR	20-OCT-2000;	2000US-0241221P.	CC	diagnosis, treatment, prevention and/or prognosis of a wide range of	
PR	20-OCT-2000;	2000US-0241785P.	CC	disorders including hyperproliferative disorders (e.g. cancer),	
PR	20-OCT-2000;	2000US-0241786P.			
PR	20-OCT-2000;	2000US-0241787P.			

CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC infectious disorders (e.g. influenza). The polynucleotides of the
CC invention can also be used in gene therapy. AAU22915-AAU23814 represent
CC the novel human enzyme polypeptides of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 452 AA;

Query Match 16.5%; Score 331; DB 4; Length 452;
Best Local Similarity 46.6%; Pred. No. 6.3e-17;
Matches 62; Conservative 20; Mismatches 29; Indels 22; Gaps 1;

[illegible]

Qy		61	PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV	120
			: :	
Db		52	PIRFWILDSPAPPICFKPFTANNGILVGKHVDAGRIYLPYLQNSHPKSVIVGLIKE	111

Qy	121	MIVVFGDEPPVFS	133
		: :	
D _b	112	MIAKFQEELPMYS	124

RESULT 25
AAG13899
ID AAG13899 standard; protein; 288 AA.
XX
XX AAG13899;
XX
DT 17-OCT-2000 (first entry)

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
XX
OS *Arabidopsis thaliana*.

25-FEB-2000; 2000EP-00301439.

PR	25-FEB-1999,	99US-0121825P.
PR	05-MAR-1999,	99US-0123180P.
PR	09-MAR-1999,	99US-0123548P.
PR	23-MAR-1999,	99US-0125788P.
PR	25-MAR-1999,	99US-0125264P.
PR	29-MAR-1999,	99US-0126785P.
PR	01-APR-1999,	99US-0127462P.
PR	06-APR-1999,	99US-0128234P.
PR	08-APR-1999,	99US-0128714P.
PR	16-APR-1999,	99US-0129845P.
PR	19-APR-1999,	99US-0130077P.
PR	21-APR-1999,	99US-0130449P.
PR	23-APR-1999,	99US-0130510P.
PR	23-APR-1999,	99US-0130891P.
PR	28-APR-1999,	99US-0131449P.
PR	30-APR-1999,	99US-0132048P.
PR	30-APR-1999,	99US-0132407P.
PR	04-MAY-1999,	99US-0132484P.
PR	05-MAY-1999,	99US-0132485P.
PR	06-MAY-1999,	99US-0132486P.

PR	06-MAY-1999;	99US-01324847P;
PR	07-MAY-1999;	99US-01328663P;
PR	11-MAY-1999;	99US-01342566P;
PR	14-MAY-1999;	99US-01342188P;
PR	14-MAY-1999;	99US-01342199P;
PR	14-MAY-1999;	99US-01342211P;
PR	14-MAY-1999;	99US-01343700P;
PR	18-MAY-1999;	99US-01347688P;
PR	19-MAY-1999;	99US-01349411P;
PR	20-MAY-1999;	99US-01351242P;
PR	21-MAY-1999;	99US-01356533P;
PR	24-MAY-1999;	99US-01356239P;
PR	25-MAY-1999;	99US-01360219P;
PR	27-MAY-1999;	99US-01363922P;
PR	28-MAY-1999;	99US-01367829P;
PR	01-JUN-1999;	99US-01372022P;
PR	03-JUN-1999;	99US-01375288P;
PR	04-JUN-1999;	99US-01375022P;
PR	07-JUN-1999;	99US-01377502P;
PR	08-JUN-1999;	99US-01380094P;
PR	10-JUN-1999;	99US-01385400P;
PR	10-JUN-1999;	99US-01388470P;
PR	14-JUN-1999;	99US-01391199P;
PR	16-JUN-1999;	99US-01394522P;
PR	16-JUN-1999;	99US-01394533P;
PR	17-JUN-1999;	99US-01394533P;
PR	18-JUN-1999;	99US-01394558P;
PR	18-JUN-1999;	99US-01394599P;
PR	18-JUN-1999;	99US-01394549P;
PR	18-JUN-1999;	99US-01394552P;
PR	18-JUN-1999;	99US-01394562P;
PR	18-JUN-1999;	99US-01394633P;
PR	18-JUN-1999;	99US-01397500P;
PR	21-JUN-1999;	99US-01397633P;
PR	21-JUN-1999;	99US-01398178P;
PR	22-JUN-1999;	99US-01398999P;
PR	23-JUN-1999;	99US-01403533P;
PR	23-JUN-1999;	99US-01403549P;
PR	24-JUN-1999;	99US-01406959P;
PR	28-JUN-1999;	99US-01408233P;
PR	30-JUN-1999;	99US-01409911P;
PR	30-JUN-1999;	99US-01428787P;
PR	01-JUL-1999;	99US-01428422P;
PR	02-JUL-1999;	99US-01421544P;
PR	02-JUL-1999;	99US-01420555P;
PR	06-JUL-1999;	99US-01423900P;
PR	08-JUL-1999;	99US-01428033P;
PR	09-JUL-1999;	99US-01429200P;
PR	12-JUL-1999;	99US-01429977P;
PR	13-JUL-1999;	99US-01435422P;
PR	14-JUL-1999;	99US-01442055P;
PR	15-JUL-1999;	99US-01443624P;
PR	16-JUL-1999;	99US-01440055P;
PR	16-JUL-1999;	99US-01440852P;
PR	16-JUL-1999;	99US-01440866P;
PR	19-JUL-1999;	99US-01443325P;
PR	19-JUL-1999;	99US-01443311P;
PR	19-JUL-1999;	99US-01443322P;
PR	19-JUL-1999;	99US-01443333P;
PR	19-JUL-1999;	99US-01443344P;
PR	19-JUL-1999;	99US-01443355P;
PR	20-JUL-1999;	99US-01443352P;
PR	20-JUL-1999;	99US-01446322P;
PR	20-JUL-1999;	99US-01448844P;
PR	21-JUL-1999;	99US-01448141P;
PR	21-JUL-1999;	99US-01450862P;
PR	22-JUL-1999;	99US-01450855P;
PR	22-JUL-1999;	99US-01450877P;
PR	22-JUL-1999;	99US-01450899P;

PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150894P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155699P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.

PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161520P.
PR 28-OCT-1999; 99US-0161522P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.4%; Score 267.5; DB 3; Length 288;
Best Local Similarity 28.6%; Pred. No. 2.7e-12;
Matches 82; Conservative 53; Mismatches 121; Indels 31; Gaps 9;

Qy 92 HVDANGKIVLPYLHEWKHPQSDLLGLIQVMVVVFGDEPPVFSRPIASYPYQATGPPNT 151
Db 9 HVTPLGLVSLPYLQNVWYFSSNLVDLSAFAFADPPLYSRRRQPPPP-----SPTV 64
Qy 152 --SYMFCMEGGISPYSGYPPNPSPGCGPYPPGYPATTSQYSPSQPPVTTVGPSTRDG 209
Db 65 YDSSLRPPSADQSLPRPPPPSPYG-----GGVSRVQVQVHVHQQQSDDAAEVFKRN 116
Qy 210 TISE--DTIRASLISAVSKLRWKEEMDRQAELNALKTEEDLKKGHKLEEMVTRL 267
Db 117 AINKVMVMVHSDLV--MRRAREAEABELLSQA---GLKRREDELNIG---LKEMV--- 165
Qy 268 DOEVAEVDKNIELKKKDEELSSALEKMNQSEN---NDIDEVIITAPILYKQILNIYAE 324
Db 166 -EKETLEQQLQIISNMNTDILDSWVRENQKTKNLVDLDVDNAFECGDTLSKQMLECTAL 224
Qy 325 ENAIEDTIFYLGEALRRGVLDLVFLKHVLLSRKQFOLRALMQKAR 371
Db 225 DLAIEDATYSLDKSFQDGVVFPQYLRNVLLSREQFFHFRATGSKVR 271

Search completed: July 12, 2004, 08:35:25
Job time : 52.4336 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:32:51 ; Search time 298.607 Seconds
(without alignments)

396.939 Million cell updates/sec

Title: US-09-804-690-4

Perfect score: 2002

Sequence: 1 MVSKYKRDLTREIVNVIT.....FOURLMQKARKTAGLSLDY 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1279676 seqs, 311918243 residues

Total number of hits satisfying chosen parameters: 1279676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2002	100.0	380	9 US-09-804-690-4	Sequence 4, Appli
2	2002	100.0	390	14 US-10-243-815A-1	Sequence 1, Appli
3	2002	100.0	390	14 US-10-053-975A-1	Sequence 1, Appli
4	2002	100.0	390	14 US-10-376-564-2	Sequence 2, Appli
5	1900.5	94.9	381	9 US-09-804-690-2	Sequence 2, Appli
6	1900.5	94.9	381	14 US-10-243-815A-3	Sequence 3, Appli
7	1900.5	94.9	391	14 US-10-205-194-109	Sequence 109, App
8	1900.5	94.9	391	14 US-10-376-564-1	Sequence 1, Appli
9	1404	70.1	285	14 US-10-376-564-82	Sequence 82, Appli
10	1404	70.1	307	15 US-10-264-049-2861	Sequence 2861, Ap
11	396.5	19.8	402	16 US-10-437-963-164064	Sequence 164064,
12	390	19.5	87	14 US-10-029-386-28907	Sequence 28907, A
13	346	17.3	146	9 US-09-925-299-1257	Sequence 1257, Ap
14	346	17.3	146	10 US-09-925-299-1257	Sequence 1257, Ap
15	328	16.4	73	14 US-10-029-386-29129	Sequence 29129, A

16	11.5	231	231	12	US-10-424-599-239306	Sequence 239306,
17	200	296	296	9	US-09-801-368-386	Sequence 386, App
18	169.5	8.5	191	12	US-10-424-599-283387	Sequence 283387,
19	161	8.0	485	9	US-09-925-300-1664	Sequence 1664, Ap
20	160	8.0	466	16	US-10-408-765A-377	Sequence 377, App
21	148.5	7.4	148	12	US-10-115-123-453	Sequence 453, App
22	148.5	7.4	148	14	US-10-012-542-453	Sequence 453, App
23	148.5	7.4	176	14	US-10-106-698-5906	Sequence 5906, Ap
24	148.5	7.4	397	15	US-10-094-749-2383	Sequence 2383, Ap
25	145	7.2	483	16	US-10-437-963-191498	Sequence 191498,
26	144.5	7.2	777	15	US-10-264-049-2971	Sequence 2971, Ap
27	142.5	7.1	587	16	US-10-437-963-106668	Sequence 106668,
28	142	7.1	505	16	US-10-408-765A-539	Sequence 539, App
29	142	7.1	983	12	US-10-112-944-435	Sequence 435, App
30	142	7.1	1021	16	US-10-408-765A-313	Sequence 313, App
31	142	7.1	1251	12	US-10-112-944-881	Sequence 881, App
32	139.5	7.0	258	12	US-10-424-599-176931	Sequence 176931,
33	139.5	7.0	278	12	US-10-425-114-68284	Sequence 68284, A
34	138.5	6.9	827	16	US-10-149-310-68	Sequence 68, Appl
35	136	6.8	1066	16	US-10-437-963-157156	Sequence 157156,
36	135.5	6.8	180	16	US-10-437-963-191496	Sequence 191496,
37	133.5	6.7	336	16	US-10-437-963-201912	Sequence 201912,
38	132	6.6	371	10	US-09-284-320-2	Sequence 2, Appli
39	132	6.6	371	14	US-10-197-666A-4	Sequence 4, Appli
40	132	6.6	371	14	US-10-024-298A-41	Sequence 41, Appl
41	132	6.6	371	14	US-10-042-211A-41	Sequence 41, Appl
42	132	6.6	371	16	US-10-617-217A-41	Sequence 41, Appl
43	131.5	6.6	605	16	US-10-437-963-102875	Sequence 102875,
44	131.5	6.6	889	16	US-10-437-963-197945	Sequence 197945,
45	131	6.5	518	15	US-10-369-493-3240	Sequence 3240, Ap
46	131	6.5	588	16	US-10-437-963-150465	Sequence 150465,
47	131	6.5	873	16	US-10-437-963-142576	Sequence 142576,
48	130.5	6.5	237	15	US-10-144-194A-30	Sequence 30, Appl
49	130.5	6.5	338	14	US-10-205-823-134	Sequence 134, App
50	130.5	6.5	1322	15	US-10-369-493-13040	Sequence 4, Appli
51	130	6.5	1647	10	US-09-824-574-4	Sequence 133141,
52	129.5	6.5	358	16	US-10-437-963-153141	Sequence 153141,
53	129.5	6.5	753	15	US-10-369-493-2333	Sequence 1810, Ap
54	129	6.4	1449	16	US-10-408-765A-1810	Sequence 1810, Ap
55	128.5	6.4	262	12	US-10-133-234A-5	Sequence 5, Appli
56	128.5	6.4	754	16	US-10-437-963-112229	Sequence 112229,
57	128	6.4	1121	16	US-10-437-963-196036	Sequence 196036,
58	127.5	6.4	284	12	US-10-206-915-236	Sequence 236, App
59	127.5	6.4	284	12	US-10-199-670-236	Sequence 236, App
60	127.5	6.4	284	12	US-10-201-858-236	Sequence 236, App
61	127.5	6.4	284	12	US-10-205-890-236	Sequence 236, App
62	127.5	6.4	284	12	US-10-208-024-236	Sequence 236, App
63	127.5	6.4	284	12	US-10-201-853-236	Sequence 236, App
64	127.5	6.4	284	12	US-10-063-745-62	Sequence 62, Appl
65	127.5	6.4	284	12	US-10-063-512-62	Sequence 62, Appl
66	127.5	6.4	284	12	US-10-063-513-62	Sequence 62, Appl
67	127.5	6.4	284	12	US-10-063-515-62	Sequence 62, Appl
68	127.5	6.4	284	12	US-10-063-549-62	Sequence 62, Appl
69	127.5	6.4	284	12	US-10-063-569-62	Sequence 62, Appl
70	127.5	6.4	284	12	US-10-063-551-62	Sequence 236, App
71	127.5	6.4	284	12	US-10-174-581-236	Sequence 236, App
72	127.5	6.4	284	12	US-10-176-483-236	Sequence 236, App
73	127.5	6.4	284	12	US-10-176-749-236	Sequence 236, App
74	127.5	6.4	284	12	US-10-176-914-236	Sequence 236, App
75	127.5	6.4	284	12	US-10-176-915-236	Sequence 236, App
76	127.5	6.4	284	12	US-10-063-555-62	Sequence 62, Appl
77	127.5	6.4	284	12	US-10-063-563-62	Sequence 62, Appl
78	127.5	6.4	284	12	US-10-063-594-62	Sequence 62, Appl
79	127.5	6.4	284	12	US-10-063-553-62	Sequence 62, Appl
80	127.5	6.4	284	12	US-10-063-554-62	Sequence 62, Appl
81	127.5	6.4	284	12	US-10-176-484-236	Sequence 236, App
82	127.5	6.4	284	12	US-10-180-550-236	Sequence 236, App
83	127.5	6.4	284	12	US-10-183-014-236	Sequence 236, App
84	127.5	6.4	284	12	US-10-187-738-236	Sequence 236, App
85	127.5	6.4	284	12	US-10-187-740-236	Sequence 236, App
86	127.5	6.4	284	12	US-10-187-883-236	Sequence 236, App
87	127.5	6.4	284	12	US-10-194-363-236	Sequence 236, App
88	127.5	6.4	284	12	US-10-194-460-236	Sequence 236, App

89 127.5 6.4 284 12 US-10-194-463-236 Sequence 236, App
90 127.5 6.4 284 12 US-10-194-484-236 Sequence 236, App
91 127.5 6.4 284 12 US-10-195-884-236 Sequence 236, App
92 127.5 6.4 284 12 US-10-195-896-236 Sequence 236, App
93 127.5 6.4 284 12 US-10-196-744-236 Sequence 236, App
94 127.5 6.4 284 12 US-10-196-755-236 Sequence 236, App
95 127.5 6.4 284 12 US-10-196-757-236 Sequence 236, App
96 127.5 6.4 284 12 US-10-197-704-236 Sequence 236, App
97 127.5 6.4 284 12 US-10-197-710-236 Sequence 236, App
98 127.5 6.4 284 12 US-10-198-758-236 Sequence 236, App
99 127.5 6.4 284 12 US-10-198-766-236 Sequence 236, App
100 127.5 6.4 284 12 US-10-199-304-236 Sequence 236, App

ALIGNMENTS

RESULT 1
US-09-804-690-4
; Sequence 4, Application US/09804690
; Patent No. US20020034743A1
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/804,690
; FILING DATE: 12-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/146,187
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-398-3249
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-804-690-4

Query Match 100.0%; Score 2002; DB 9; Length 380;
Best Local Similarity 100.0%; Pred. No. 4.7e-139; Indels 0; Gaps 0;
Matches 380; Conservative 0; Mismatches 0;

Qy 1 MVSQKYRDLTVRETNVITLYKDLKPVLDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
Db 1 MVSQKYRDLTVRETNVITLYKDLKPVLDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
Qy 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHEWKHPQSDLLGLTQV 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHEWKHPQSDLLGLTQV 120

Qy 121 MIVVFGDEPPVFSRPIASYPYQATGPNTSYMPGPGGISPYPSGYPPNPSGYPGCPY 180
Db 121 MIVVFGDEPPVFSRPIASYPYQATGPNTSYMPGPGGISPYPSGYPPNPSGYPGCPY 180
Qy 181 PPGGYPYPATTSQYPSQPPVTTVGPSSRDGTISEDITIRASLISAVSDKLRWKKEMDMRAQ 240
Db 181 PPGGYPYPATTSQYPSQPPVTTVGPSSRDGTISEDITIRASLISAVSDKLRWKKEMDMRAQ 240
Qy 241 AELNALKRTBEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKEMNQSE 300
Db 241 AELNALKRTBEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKEMNQSE 300
Qy 301 NNDIDEVIIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDVFLKHVRLLSRKQ 360
Db 301 NNDIDEVIIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDVFLKHVRLLSRKQ 360
Qy 361 FQLRALMQKARKTAGLSLDLY 380
Db 361 FQLRALMQKARKTAGLSLDLY 380

RESULT 2

US-10-243-815A-1
; Sequence 1, Application US/10243815A
; Publication No. US2003009988A1
; GENERAL INFORMATION:
; APPLICANT: La Brie, Samuel T.; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR SUPPRESSOR
; FILE REFERENCE: PF-0199-2 DIV
; CURRENT APPLICATION NUMBER: US/10/243,815A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 09/216,387
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: US 08/786,999
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2003009988A1 609476CD1
US-10-243-815A-1

Query Match 100.0%; Score 2002; DB 14; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.9e-139; Indels 0; Gaps 0;
Matches 380; Conservative 0; Mismatches 0;

Qy 1 MVSQKYRDLTVRETNVITLYKDLKPVLDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 60
Db 11 MVSQKYRDLTVRETNVITLYKDLKPVLDSYVFNDSGSSRELMLNLTGTIPVYRGNTYNI 70
Qy 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHEWKHPQSDLLGLTQV 120
Db 71 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHEWKHPQSDLLGLTQV 130
Qy 121 MIVVFGDEPPVFSRPIASYPYQATGPNTSYMPGPGGISPYPSGYPPNPSGYPGCPY 180
Db 131 MIVVFGDEPPVFSRPIASYPYQATGPNTSYMPGPGGISPYPSGYPPNPSGYPGCPY 190
Qy 181 PPGGYPYPATTSQYPSQPPVTTVGPSSRDGTISEDITIRASLISAVSDKLRWKKEMDMRAQ 240
Db 191 PPGGYPYPATTSQYPSQPPVTTVGPSSRDGTISEDITIRASLISAVSDKLRWKKEMDMRAQ 250
Qy 241 AELNALKRTBEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKEMNQSE 300
Db 251 AELNALKRTBEDLKKGHOKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEKEMNQSE 310
Qy 301 NNDIDEVIIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDVFLKHVRLLSRKQ 360
Db 301 NNDIDEVIIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDVFLKHVRLLSRKQ 360

```
Db 311 NNDIDEVIPTAPLYKQILNLVAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 370
QY 361 FQRLMOKARKTAGLSLDLY 380
Db 371 FQRLMOKARKTAGLSLDLY 390

RESULT 3
US-10-053-975A-1
; Sequence 1, Application US/10053975A
; Publication No. US20030138639A1
; GENERAL INFORMATION:
; APPLICANT: LI, LIMIN
; APPLICANT: COHEN, STANLEY N.
; TITLE OF INVENTION: "Mammalian Tumor Susceptibility Gene
; TITLE OF INVENTION: Products and Their Uses"
; FILE REFERENCE: STAN-216
; CURRENT APPLICATION NUMBER: US/10/053,975A
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,763
; PRIOR FILING DATE: 2001-01-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-053-975A-1

Query Match 100.0%; Score 2002; DB 14; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.9e-139;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSXYKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 60
Db 11 MVSXYKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 70
QY 61 PICWLDDTYPNPDPICFVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGLIQV 120
Db 71 PICWLDDTYPNPDPICFVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGLIQV 130
QY 121 MIVVFGDEPPVFSRPSISASYPYQATGPPNTSYMPGMPGGISPYSPGYPPNPSGYPGCPY 180
Db 131 MIVVFGDEPPVFSRPSISASYPYQATGPPNTSYMPGMPGGISPYSPGYPPNPSGYPGCPY 190
QY 181 PPGGYPATTSQYPSQPPVTVVTPGSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
Db 191 PPGGYPATTSQYPSQPPVTVVTPGSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250
QY 241 AELNALKRTEDDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDEELSSALEKMNQSE 300
Db 251 AELNALKRTEDDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDEELSSALEKMNQSE 310
QY 301 NNDIDEVIPTAPLYKQILNLVAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 360
Db 311 NNDIDEVIPTAPLYKQILNLVAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 370
QY 361 FQRLMOKARKTAGLSLDLY 380
Db 371 FQRLMOKARKTAGLSLDLY 390

RESULT 4
US-10-376-564-2
; Sequence 2, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jörn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
```

```
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; FILE REFERENCE: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-376-564-2

Query Match 100.0%; Score 2002; DB 14; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.9e-139;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSXYKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 60
Db 11 MVSXYKYRDLTVRETNNVITLYKDLKPVLDVSVFNDGSSRELMLNLTGTIPVYRGNTYNI 70
QY 61 PICWLDDTYPNPDPICFVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGLIQV 120
Db 71 PICWLDDTYPNPDPICFVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGLIQV 130
QY 121 MIVVFGDEPPVFSRPSISASYPYQATGPPNTSYMPGMPGGISPYSPGYPPNPSGYPGCPY 180
Db 131 MIVVFGDEPPVFSRPSISASYPYQATGPPNTSYMPGMPGGISPYSPGYPPNPSGYPGCPY 190
QY 181 PPGGYPATTSQYPSQPPVTVVTPGSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
Db 191 PPGGYPATTSQYPSQPPVTVVTPGSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250
QY 241 AELNALKRTEDDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDEELSSALEKMNQSE 300
Db 251 AELNALKRTEDDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDEELSSALEKMNQSE 310
QY 301 NNDIDEVIPTAPLYKQILNLVAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 360
Db 311 NNDIDEVIPTAPLYKQILNLVAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 370
QY 361 FQRLMOKARKTAGLSLDLY 380
Db 371 FQRLMOKARKTAGLSLDLY 390

RESULT 5
US-09-804-690-2
; Sequence 2, Application US/09804690
; Patent No. US20020034743A1
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/804,690
; FILING DATE: 12-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/146,187
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-804-690-2

Query Match          94.9%; Score 1900.5; DB 9; Length 381;
Best Local Similarity 94.5%; Pred. No. 1.4e-131;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY      1  MMSKYKYRDLTVRETNVITLYKDLKPVLDVSVFNDGSSRELNMNLTGTIPVYRGNTYNI 60
DB      1  MMSKYKYRDLTVRQTVNVNVIAMKYKDLKPVLDVSVFNDGSSRELNMNLTGTIPVYRGNTYNI 60
QY      61  PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
DB      61  PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
QY      121  MIWFGDEPPVFSRP-ISASYPYQATGPNTSYMPGPGGISPYPSPGPNPSPGPGCP 179
DB      121  MIWFGDEPPVFSRP-ISASYPYQATGPNTSYMPGPGGISPYPSPGPNPSPGPGCP 180
QY      180  YPGGYPATTSSQYPSQPPVTVGPRDGTISEDITRASLISAVSDKLRWRMKEEMDRA 239
DB      181  YPPAGYPATTSSQYPSQPPVTVGPRDGTISEDITRASLISAVSDKLRWRMKEEMDGA 240
QY      240  QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDEELSSALEKMNQS 299
DB      241  QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDEELSSALEKMNQS 300
QY      300  ENNDIDEVIPTAPLYKQILNLYAEENAIETDITFYLGEALRRGVLDLDFLKHVRLLSRK 359
DB      301  ENNDIDEVIPTAPLYKQILNLYAEENAIETDITFYLGEALRRGVLDLDFLKHVRLLSRK 360
QY      360  QFQRLALMOKARKTAGLSLDLY 380
DB      361  QFQRLALMOKARKTAGLSLDLY 381

RESULT 6
US-10-243-815A-3
; Sequence 3, Application US/10243815A
; Publication No. US2003009988A1
; GENERAL INFORMATION:
; APPLICANT: La Brie, Samuel T.; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR SUPPRESSOR
; FILE REFERENCE: PF-0199-2 DIV
; CURRENT APPLICATION NUMBER: US/10/243,815A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 09/216,387
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: US 08/786,999
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank ID No. US2003009988A1 gi330330
US-10-243-815A-3

Query Match          94.9%; Score 1900.5; DB 14; Length 381;
Best Local Similarity 94.5%; Pred. No. 1.4e-131;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY      1  MMSKYKYRDLTVRETNVITLYKDLKPVLDVSVFNDGSSRELNMNLTGTIPVYRGNTYNI 60
DB      1  MMSKYKYRDLTVRQTVNVNVIAMKYKDLKPVLDVSVFNDGSSRELNMNLTGTIPVYRGNTYNI 60
QY      61  PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
DB      61  PICLWLLDTPYNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
QY      121  MIWFGDEPPVFSRP-ISASYPYQATGPNTSYMPGPGGISPYPSPGPNPSPGPGCP 179
DB      121  MIWFGDEPPVFSRP-ISASYPYQATGPNTSYMPGPGGISPYPSPGPNPSPGPGCP 180
QY      180  YPGGYPATTSSQYPSQPPVTVGPRDGTISEDITRASLISAVSDKLRWRMKEEMDRA 239
DB      181  YPPAGYPATTSSQYPSQPPVTVGPRDGTISEDITRASLISAVSDKLRWRMKEEMDGA 240
QY      240  QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDEELSSALEKMNQS 299
DB      241  QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLLKKKDEELSSALEKMNQS 300
QY      300  ENNDIDEVIPTAPLYKQILNLYAEENAIETDITFYLGEALRRGVLDLDFLKHVRLLSRK 359
DB      301  ENNDIDEVIPTAPLYKQILNLYAEENAIETDITFYLGEALRRGVLDLDFLKHVRLLSRK 360
QY      360  QFQRLALMOKARKTAGLSLDLY 380
DB      361  QFQRLALMOKARKTAGLSLDLY 381

RESULT 7
US-10-205-194-109
; Sequence 109, Application US/10205194
; Publication No. US20030134301A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brookesbank, Robert
; APPLICANT: Pimock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018201
; CURRENT APPLICATION NUMBER: US/10/205,194
; CURRENT FILING DATE: 5200-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Tumor susceptibility protein TSG101
US-10-205-194-109

Query Match          94.9%; Score 1900.5; DB 14; Length 391;
Best Local Similarity 94.5%; Pred. No. 1.4e-131;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
```

```
QY 1 MVSXKYRDLTVRETWNVITLYKDLKPVLDVSYFVNDGSSRELNMLTGTTPVYRGNTYNI 60
Db 11 MMSXKYRDLTVRETWNVITLYKDLKPVLDVSYFVNDGSSRELNMLTGTTPVYRGNTYNI 70
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHVDANGKIYLYPLHWHKHPQSDLLGLIQV 120
Db 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHVDANGKIYLYPLHWHKHPQSDLLGLIQI 130
QY 121 MIVVFGDEPPVFSRPI-ISASYPYQATGPNTSYMPGMPGSGISAYPSGYPPNPSGYGCP 179
Db 131 MIVVFGDEPPVFSRPI-ISASYPYQATGPNTSYMPGMPGSGISAYPSGYPPNPSGYGCP 190
QY 180 YPPGGPYPATTSQYPSQPPVTTVGSRDGTISEDTIRASLIYSAVSDKLRWRMKEMDRA 239
Db 191 YPPAGPYPATTSQYPSQPPVTTVGSRDGTISEDTIRASLIYSAVSDKLRWRMKEMDGA 250
QY 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVVDKNIELLKKKDEELSSALEKMNOS 299
Db 251 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVVDKNIELLKKKDEELSSALEKMNOS 310
QY 300 ENNDIDEVIITPAPLYKQILNLYABENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 359
Db 311 ENNDIDEVIITPAPLYKQILNLYABENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 370
QY 360 QFQLRALMOKARKTAGLSLDLY 380
Db 371 QFQLRALMOKARKTAGLSLDLY 391
RESULT 8
US-10-376-564-1
; Sequence 1, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jörn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-376-564-1
Query Match 94.9%; Score 1900.5; DB 14; Length 391;
Best Local Similarity 94.5%; Pred. No. 1.4e-131;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MVSXKYRDLTVRETWNVITLYKDLKPVLDVSYFVNDGSSRELNMLTGTTPVYRGNTYNI 60
Db 11 MMSXKYRDLTVRETWNVITLYKDLKPVLDVSYFVNDGSSRELNMLTGTTPVYRGNTYNI 70
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHVDANGKIYLYPLHWHKHPQSDLLGLIQV 120
Db 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHVDANGKIYLYPLHWHKHPQSDLLGLIQI 130
QY 121 MIVVFGDEPPVFSRPI-ISASYPYQATGPNTSYMPGMPGSGISAYPSGYPPNPSGYGCP 179
```

```
Db 131 MIVVFGDEPPVFSRPI-ISASYPYQATGPNTSYMPGMPGSGISAYPSGYPPNPSGYGCP 190
QY 180 YPPGGPYPATTSQYPSQPPVTTVGSRDGTISEDTIRASLIYSAVSDKLRWRMKEMDRA 239
Db 191 YPPAGPYPATTSQYPSQPPVTTVGSRDGTISEDTIRASLIYSAVSDKLRWRMKEMDGA 250
QY 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVVDKNIELLKKKDEELSSALEKMNOS 299
Db 251 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVVDKNIELLKKKDEELSSALEKMNOS 310
QY 300 ENNDIDEVIITPAPLYKQILNLYABENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 359
Db 311 ENNDIDEVIITPAPLYKQILNLYABENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 370
QY 360 QFQLRALMOKARKTAGLSLDLY 380
Db 371 QFQLRALMOKARKTAGLSLDLY 391
RESULT 9
US-10-376-564-82
; Sequence 82, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jörn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-564-82
Query Match 70.1%; Score 1404; DB 14; Length 285;
Best Local Similarity 100.0%; Pred. No. 3.3e-95;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 110 QSDLLGLIQVMIIVVFGDEPPVFSRPI-ISASYPYQATGPNTSYMPGMPGSGISAYPSGY 169
Db 15 QSDLLGLIQVMIIVVFGDEPPVFSRPI-ISASYPYQATGPNTSYMPGMPGSGISAYPSGY 74
QY 170 PMSGYGCPYPPGPGYPATTSQYPSQPPVTTVGSRDGTISEDTIRASLIYSAVSDKLR 229
Db 75 PMSGYGCPYPPGPGYPATTSQYPSQPPVTTVGSRDGTISEDTIRASLIYSAVSDKLR 134
QY 230 WRKKEEMDRAQAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVVDKNIELLKKKDEELS 289
Db 135 WRKKEEMDRAQAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVVDKNIELLKKKDEELS 194
QY 290 SALEKMNQSENNDIDEVIITPAPLYKQILNLYABENAIEDTIFYLGEALRRGVLDLDFV 349
Db 195 SALEKMNQSENNDIDEVIITPAPLYKQILNLYABENAIEDTIFYLGEALRRGVLDLDFV 254
QY 350 LKHVRLLSRKQFQLRALMOKARKTAGLSLDLY 380
```

```
Db      255 LKHVLLSRKQFQRLALMOKARKTAGLSLDLY 285
;
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 164064
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_62C.1.pap
; US-10-437-963-164064

Query Match          19.8%; Score 396.5; DB 16; Length 402;
Best Local Similarity 30.2%; Pred. No. 7.3e-21;
Matches 108; Conservative 66; Mismatches 147; Indels 37; Gaps 10;

QY      22 YKDLKPVLDYVFNDSRELMLNLTCTIPVVRGNTYNTIPICLWLDTYYPNPPICFVKP 81
;      :      :      :      :      :      :      :      :      :      :
Db      56 FPSLHPEKALFTHNDGRAHLLQADGTIPHHAGASYNLPAVLWLPPEYPRSPPLVFLSP 115
;      :      :      :      :      :      :      :      :      :      :
QY      82 TSSMTIKTKGH--VDANGKI-YLPYLHEMKHPQSDLLGLIQVMIVVFGDEPPVFSRPIA 138
;      :      :      :      :      :      :      :      :      :      :
Db      116 TRDMVTKP-HHPLVDRSGLVANAPYLRSWFPGSNLVDLRSLSHLFGLDPLFTR---- 170
;      :      :      :      :      :      :      :      :      :      :
QY      139 SYPPYQATGPPNTSYMPGMPGGISYPSPGYPNPSPGYPGCGYPGYPATTSQYPSQP 198
;      :      :      :      :      :      :      :      :      :      :
Db      171 -----SPNPPPPSPSPPIPA--TLPFRVHPSSSS-----PSPSPYRPPASPQLAARP 216
;      :      :      :      :      :      :      :      :      :      :
QY      199 PVTTVGPRSD-GTISEDTRASLIIS---AVSDKLRWRMKEEMDRACAEALNALKRTBEDLK 254
;      :      :      :      :      :      :      :      :      :      :
Db      217 P-----PTEDPAEYVYKRNALAKLVDMAADATLRPREAEVDTLFAMQATLRSRGVVS 271
;      :      :      :      :      :      :      :      :      :      :
QY      255 KGHOKLEEMVTRLDQEAEDVKNIELLKKDEELSSALEKMNQSENNDIDEVIITAPL 314
;      :      :      :      :      :      :      :      :      :      :
Db      272 DGVKKMGEEKEALERRLDQVMATDLMW-----AWMENTKGAAGDTEADEALETADVL 325
;      :      :      :      :      :      :      :      :      :      :
QY      315 YKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFVLKHVLLSRKQFQRLALMOKARK 372
;      :      :      :      :      :      :      :      :      :      :
Db      326 SKQMLECTAADLALEDITVADLKAIQBGSVFDFGYLRSVRALAREQFQFVLTSTKVNK 383
;      :      :      :      :      :      :      :      :      :      :

RESULT 12
US-10-029-386-28907
; Sequence 28907, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28907
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR11.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.83
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.6
; OTHER INFORMATION: SWISSPROT HIT: Q99816, EVALUATION 3.00e-38
; US-10-029-386-28907

Query Match          19.5%; Score 390; DB 14; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.9e-21;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
US-10-029-386-28907
```

```
Db      255 LKHVLLSRKQFQRLALMOKARKTAGLSLDLY 285
;
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 164064
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_62C.1.pap
; US-10-437-963-164064

Query Match          19.8%; Score 396.5; DB 16; Length 402;
Best Local Similarity 30.2%; Pred. No. 7.3e-21;
Matches 108; Conservative 66; Mismatches 147; Indels 37; Gaps 10;

QY      22 YKDLKPVLDYVFNDSRELMLNLTCTIPVVRGNTYNTIPICLWLDTYYPNPPICFVKP 81
;      :      :      :      :      :      :      :      :      :      :
Db      56 FPSLHPEKALFTHNDGRAHLLQADGTIPHHAGASYNLPAVLWLPPEYPRSPPLVFLSP 115
;      :      :      :      :      :      :      :      :      :      :
QY      82 TSSMTIKTKGH--VDANGKI-YLPYLHEMKHPQSDLLGLIQVMIVVFGDEPPVFSRPIA 138
;      :      :      :      :      :      :      :      :      :      :
Db      116 TRDMVTKP-HHPLVDRSGLVANAPYLRSWFPGSNLVDLRSLSHLFGLDPLFTR---- 170
;      :      :      :      :      :      :      :      :      :      :
QY      139 SYPPYQATGPPNTSYMPGMPGGISYPSPGYPNPSPGYPGCGYPGYPATTSQYPSQP 198
;      :      :      :      :      :      :      :      :      :      :
Db      171 -----SPNPPPPSPSPPIPA--TLPFRVHPSSSS-----PSPSPYRPPASPQLAARP 216
;      :      :      :      :      :      :      :      :      :      :
QY      199 PVTTVGPRSD-GTISEDTRASLIIS---AVSDKLRWRMKEEMDRACAEALNALKRTBEDLK 254
;      :      :      :      :      :      :      :      :      :      :
Db      217 P-----PTEDPAEYVYKRNALAKLVDMAADATLRPREAEVDTLFAMQATLRSRGVVS 271
;      :      :      :      :      :      :      :      :      :      :
QY      255 KGHOKLEEMVTRLDQEAEDVKNIELLKKDEELSSALEKMNQSENNDIDEVIITAPL 314
;      :      :      :      :      :      :      :      :      :      :
Db      272 DGVKKMGEEKEALERRLDQVMATDLMW-----AWMENTKGAAGDTEADEALETADVL 325
;      :      :      :      :      :      :      :      :      :      :
QY      315 YKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFVLKHVLLSRKQFQRLALMOKARK 372
;      :      :      :      :      :      :      :      :      :      :
Db      326 SKQMLECTAADLALEDITVADLKAIQBGSVFDFGYLRSVRALAREQFQFVLTSTKVNK 383
;      :      :      :      :      :      :      :      :      :      :

RESULT 12
US-10-029-386-28907
; Sequence 28907, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28907
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR11.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.83
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.6
; OTHER INFORMATION: SWISSPROT HIT: Q99816, EVALUATION 3.00e-38
; US-10-029-386-28907

Query Match          19.5%; Score 390; DB 14; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.9e-21;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
US-10-029-386-28907
```

QY 272 AAVDKNIILLKKKDELSALEKMNQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDT 331
Db 9 AAVDKNIILLKKKDELSALEKMNQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDT 68
QY 332 IFYLGALRRRGVIDLDVFL 350
Db 69 IFYLGALRRRGVIDLDVFL 87

RESULT 13

US-09-925-299-1257
; Sequence 1257, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1257
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (131)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (135)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (138)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1257

Query Match 17.3%; Score 346; DB 9; Length 146;
Best Local Similarity 95.5%; Pred. No. 1e-17;
Matches 63; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 33 VFNDGSSRELNLGTIPVVRGNTYNIPICLWLDTPYNNPPICFVKPTSSMTIKTKH 92
Db 69 VFNDGSSRELNLGTIPVVRGNTYNIPICLWLDTPYNNPPICFVKPTSSMTIKTKH 128
QY 93 VDANGK 98
Db 129 VDXPKK 134

RESULT 14

US-09-925-299-1257
; Sequence 1257, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1257
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (131)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (135)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (138)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1257

Query Match 17.3%; Score 346; DB 10; Length 146;
Best Local Similarity 95.5%; Pred. No. 1e-17;
Matches 63; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 33 VFNDGSSRELNLGTIPVVRGNTYNIPICLWLDTPYNNPPICFVKPTSSMTIKTKH 92
Db 69 VFNDGSSRELNLGTIPVVRGNTYNIPICLWLDTPYNNPPICFVKPTSSMTIKTKH 128
QY 93 VDANGK 98
Db 129 VDXPKK 134

RESULT 15

US-10-029-386-29129
; Sequence 29129, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENOTYPING
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29129
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR11.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2

```
;
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: Q99816, EVALUE 4.00e-31
US-10-029-386-29129

Query Match      16.4%; Score 328; DB 14; Length 73;
Best Local Similarity 84.8%; Pred. No. 8.4e-17; Indels 0; Gaps 0;
Matches 56; Conservative 3; Mismatches 7;

QY 44 NLGTGTPVVRGNTYNIPICLWLLDTYPVNPPICFVKPTSSMTIKTKGHVDANGKIYLPY 103
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
5 NRVTESISFGSNTYNIPICLWLLDTYPVNPPICFVKPTSSMTIKTKGHVDANGKIYLPY 64

QY 104 LHEWKH 109
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
65 LHEWKH 70

RESULT 16
US-10-424-599-239306
; Sequence 239306, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 239306
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58119C.1.pep
US-10-424-599-239306

Query Match      11.5%; Score 231; DB 12; Length 231;
Best Local Similarity 29.1%; Pred. No. 5.3e-09;
Matches 62; Conservative 26; Mismatches 77; Indels 48; Gaps 7;

QY 16 VNVITLYKDLKPVLDYVFDGSSRELNMNLTGTIPVYRGNTYNIPICLWLLDTYPVNP 75
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
39 VALTTAPPSLEPKTASTHNDGRSVNLLQADGTIPMTFGQVTVNIPVVIWLMESYPRHP 98

QY 76 ICFVKPTSSMTIK-TGKHVDANGKIYLYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSR 134
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
99 CVYVNPTRDMTIKRPHPVNPGLVSPVYLQNTWYTP-----GDDSVPLQR 143

QY 135 PISASYPYQATGP-----PNTSY-----NPGMPGGISPYPSGYPPN----- 171
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
144 P-----FAPRLRGLLGRHVHFLAPRRSHHCLPFPFRATPSTHNDGRSVNLLQADGT 199

QY 172 -PSGYPCYPGPGGYPATTSQYSPQPVTV 203
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
200 IPMTFGQVTV-----NVPWISPSPCAPSSRV 228

RESULT 17
US-09-801-368-386
; Sequence 386, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
```

```
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 386
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-09-801-368-386

Query Match      10.0%; Score 200; DB 9; Length 296;
Best Local Similarity 24.2%; Pred. No. 1.4e-06;
Matches 72; Conservative 57; Mismatches 124; Indels 44; Gaps 13;

QY 7 YRD--LTVRETNNVITLYKDLKPVLDYVFDGSSRELNMNLTGTIPVYRGNT-YNIPIC 63
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
25 YNDGRTTFHDSLALLDNFHSRLRPRTRVFTHSDGTQQLLSIYGTISTGEDSSPSHPVI 84

QY 64 LWLLDTYPVNPPICT-----FVKPTSSMTIKTKGHVDANGKIYLYLHEWKHPQSDLLGL 117
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
85 MWVPSMYPVKPPFISINLENFDMNTISSLPIDQYIDNSGWIAPILKCHWDPAAMNLMV 144

QY 118 IQVMIVVFGDEPPVFSRPIASASYPYQATGPPNTSYMPGPGGISPYPSGYPPNPSGYPG 177
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
145 VQELMSLL-HEPFDQAPSLPPKPNLTQLOEQNTPLPPKPKS-----PHLKPPLP----- 194

QY 178 CYPYPGGYPATTSQYSPQPVTVVGPSSRD-----GTISEDITRASLISAVSDKLRW 230
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
195 ---PPPPPPQASNALDLMMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVADKILT 249

QY 231 R--MKEEMDRAQAEINALKRTEEDLKKGHOKLEEMVTRLDQEAQVYDKNIELKKK 284
Db | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
250 RQVVMGESIARFH-EIIA-----IDKNHLRAVEQA--LEQTMHSLNAQIDVLNRK 296

RESULT 18
US-10-424-599-283387
; Sequence 283387, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 283387
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(191)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
```


; PRIOR APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-115-123-453

Query Match
Best Local Similarity 7.4%; Score 148.5; DB 12; Length 148;
Matches 41; Conservative 18; Mismatches 46; Indels 33; Gaps 6;

QY 126 GDEPPVFSRPIASYPYQATGTPNTSYMPGMPGGISPYSGYPPNP---SGYPGC--- 178
Db 22 GGSNPAHPPIINPPFPGPCPP-----PGAHGNPAFPGGPHVPPQPGYGCQPLG 75
QY 179 PYPPGGPYPATTSSQYSPQPTVTGVRDGTISEDITIRASLISAVSDKLRWRMKEEMDR 238
Db 76 PYPPPPPPPA-----PGIPPVNPLAPGMVGP-----AVIVDK---KMQKMKK 115
QY 239 AQAEINALKRTEEDLKKG 256
Db 116 AHKKMKHKQKHKKYKHG 133

; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 2001-12-12
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-06-16
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-06-16
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-06-16
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-06-22
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-542-453

Query Match
Best Local Similarity 7.4%; Score 148.5; DB 14; Length 148;
Matches 41; Conservative 18; Mismatches 46; Indels 33; Gaps 6;

QY 126 GDEPPVFSRPIASYPYQATGTPNTSYMPGMPGGISPYSGYPPNP---SGYPGC--- 178
Db 22 GGSNPAHPPIINPPFPGPCPP-----PGAHGNPAFPGGPHVPPQPGYGCQPLG 75
QY 179 PYPPGGPYPATTSSQYSPQPTVTGVRDGTISEDITIRASLISAVSDKLRWRMKEEMDR 238
Db 76 PYPPPPPPPA-----PGIPPVNPLAPGMVGP-----AVIVDK---KMQKMKK 115
QY 239 AQAEINALKRTEEDLKKG 256
Db 116 AHKKMKHKQKHKKYKHG 133

RESULT 23
US-10-106-698-5906
; Sequence 5906, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5906
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5906

Query Match
Best Local Similarity 7.4%; Score 148.5; DB 14; Length 176;
Matches 41; Conservative 18; Mismatches 46; Indels 33; Gaps 6;

QY 126 GDEPPVFSRPIASYPYQATGTPNTSYMPGMPGGISPYSGYPPNP---SGYPGC--- 178
Db 50 GGSNPAHPPIINPPFPGPCPP-----PGAHGNPAFPGGPHVPPQPGYGCQPLG 103
QY 179 PYPPGGPYPATTSSQYSPQPTVTGVRDGTISEDITIRASLISAVSDKLRWRMKEEMDR 238
Db 104 PYPPPPPPPA-----PGIPPVNPLAPGMVGP-----AVIVDK---KMQKMKK 143
QY 239 AQAEINALKRTEEDLKKG 256
Db 144 AHKKMKHKQKHKKYKHG 161

RESULT 24
US-10-094-749-2983
; Sequence 2983, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO

```
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094, 749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350, 435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2983
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2983

Query Match      7.4%; Score 148.5; DB 15; Length 397;
Best Local Similarity 22.7%; Pred. No. 0.013;
Matches 84; Conservative 59; Mismatches 130; Indels 97; Gaps 17;

QY 51 VPYRGNTYNIPICLWLLDTYPYNPPICFVKP--TSSMTIKTKGHVDANGKIYLPYLHEWK 108
Db 54 LPFTNNITINIIILLPQFQOEKPVISYPPIRHHLMKQGVVTS-----PLVNNFT 107

QY 109 HPQSDLLGLIQMIVVFGDEPPVFSRPISASYPYQATGPPNTSYMGMPGGISYPVS-G 167
Db 108 -MHSDLGKIIQSLLEDFWKNPPVLA-PTSTAPP-----YYSNPSSGMSFYASQG 154

QY 168 YPPNPSGYGCPYPP-----GGPYPATTSQYPSQPPVTVGPPSRDGTISED 215
Db 155 PFPLP-----PYPQEANRSITSLSVADTVSSSTTSHHTAKPAAPSEGVLSNLPPIPT 208

QY 216 IRASLISAVSKLRWMEKMDR-----AGAEINAKRTEEDLKKGHKLSEMT--R 266
Db 209 VDAS--IPTSQNGFGYKMPDVPDAFFPELSELSVSQLTDMNEQEVLE----LEQFLTLQP 261

QY 267 LDQVEAEVD---KNIELLKXKDEELSSALE-----KMNQSEN 301
Db 262 LKQIITDKDLVKSIEELARKNLLLEPSLEAKRQVLDKYELLTQMKSTFERKMQORHEL 321

QY 302 NDIDEVIITAPLYKQIILNVAEENAIETIPLYLGEALRGRVIDLDVFLK-----H 352
Db 322 SE-----SCSASALQARLKVAHE--ABEESDNIAEDFLEGMKMEIDDFLSSFMEKRTICH 374

QY 353 VRLLSRKQKQ 362
Db 375 CRRAKEELQ 384

RESULT 25
US-10-437-963-191498
; Sequence 191498, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191498
```

```
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(483)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MET4530_87810C.1.pap
US-10-437-963-191498

Query Match      7.2%; Score 145; DB 16; Length 483;
Best Local Similarity 27.3%; Pred. No. 0.03;
Matches 51; Conservative 2; Mismatches 56; Indels 78; Gaps 8;

QY 50 PVPYRGNTYNIPICLWLLDTYPYNPPICFVKETSSMTIKTKGHVDANGKIYLPYLHEWK 109
Db 90 PPPYFGXPYPP-----PYPPP-----YYPYPPXHR 115

QY 110 PQSDLLGLIQMIVVFGDEPPVFSRPISASYPYQATGPPNTS-----YMGMPGGIS 162
Db 116 PTSD-----PRPRQPP-----PCPPXPXPPXPPPPPPPPPPPPPPPPPPG--X 159

QY 163 YPPSGY-----PPNPSGYGCPYPP-----GGPYPATTSQYPSQPPVT 201
Db 160 PYPYPPYLPYPPPPXSPSPYGGWYPPYPPYPPYPPYPPYAGRPYPPYPPYPPX 219

QY 202 TVGSRD 208
Db 220 IHLPTSD 226

Search completed: July 12, 2004, 08:52:19
Job time : 299.607 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 16.9777 Seconds

(without alignments)

2152.993 Million cell updates/sec

Title: US-09-804-690-4

Perfect score: 2002

Sequence: 1 MVSXKYRDLIVRETVNVIT.....FQLRALMOKARKTAGLSDLV 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	658	32.9	404	2 H88650	protein C09G12.9 [
2	388	19.4	83	2 148283	gene CC2 protein -
3	153.5	7.7	169	2 T34520	hypothetical prote
4	147	7.3	488	1 LUH07	annexin VII, long
5	145.5	7.3	212	2 S74288	hypothetical prote
6	142	7.1	505	2 A53152	annexin XI - human
7	141	7.0	463	2 S29170	annexin VII - mous
8	138.5	6.9	827	2 T39608	zinc finger transc
9	136.5	6.8	503	1 LURB11	annexin XI - rabbi
10	135.5	6.8	437	2 T14192	extensin homolog T
11	134	6.7	198	2 D70509	hypothetical prote
12	134	6.7	485	2 T37550	hypothetical colle
13	132.5	6.6	671	2 T36037	probable export as
14	131.5	6.6	370	2 T42532	hypothetical prote
15	131	6.5	1613	2 S39059	protein BRG1 - hum
16	131	6.5	1880	2 T18531	tractin - medicina
17	130.5	6.5	338	2 I53043	transforming prote
18	130	6.5	1647	2 S45252	SNF2beta protein -
19	128.5	6.4	1006	2 T42731	atrophin-1 related
20	128	6.4	139	2 H84809	hypothetical prote
21	128	6.4	3942	2 T42730	Bassoon protein -
22	127	6.3	1184	2 G01763	atrophin-1 - human
23	126.5	6.3	678	2 H88187	protein Cl8H9.8 [i
24	126	6.3	1453	2 S21626	collagen alpha 1(I
25	125	6.2	2715	2 T13049	eyelid - fruit fly
26	124.5	6.2	669	2 T28754	hypothetical prote
27	124.5	6.2	990	2 T14756	hypothetical prote
28	124	6.2	964	2 T21865	hypothetical prote
29	123.5	6.2	279	2 T05421	hypothetical prote

30	123.5	6.2	380	2 S51797	vasodilator-stimul
31	123	6.1	564	2 H70804	hypothetical prote
32	122	6.1	887	1 S57219	1-phosphatidylinos
33	122	6.1	978	2 A70387	conserved hypothet
34	122	6.1	1181	2 C86349	F8K7.4 protein - A
35	121.5	6.1	551	2 S57447	HPBR1-7 protein -
36	121	6.0	792	2 T49989	hypothetical prote
37	120.5	6.0	1902	2 C97702	cell surface antig
38	119.5	6.0	503	1 LUB011	annexin XI form A
39	119.5	6.0	505	1 S23447	annexin XI form B
40	119	5.9	577	2 T09024	proline-rich prote
41	119	5.9	1357	2 T29265	hypothetical prote
42	119	5.9	1422	2 T24212	hypothetical prote
43	118.5	5.9	384	2 S51796	vasodilator-stimul
44	118.5	5.9	400	2 E70318	hypothetical prote
45	118.5	5.9	915	2 T12526	hypothetical prote
46	118.5	5.9	1255	2 T31065	diaphanous protein
47	118.5	5.9	1281	2 JC5368	dynactin 1 - mouse
48	118	5.9	176	2 A86441	hypothetical prote
49	118	5.9	214	2 T07954	extensin-like cell
50	118	5.9	214	2 T09854	proline-rich cell
51	118	5.9	530	2 T48627	hypothetical prote
52	118	5.9	1018	2 T43168	hypothetical prote
53	118	5.9	1171	2 T17454	diaphanous-related
54	118	5.9	1259	2 T16038	hypothetical prote
55	118	5.9	1736	2 T00391	hypothetical prote
56	117.5	5.9	316	2 T20497	hypothetical prote
57	117.5	5.9	620	2 S06733	hydroxyproline-ric
58	117.5	5.9	853	2 T51505	hypothetical prote
59	117.5	5.9	1802	2 H88444	protein C26E6.12 [
60	117	5.8	177	2 S65780	glycine/proline-ri
61	117	5.8	491	2 S14182	DNA-directed RNA p
62	117	5.8	505	2 S72273	actin-depolymerizi
63	117	5.8	650	2 S14181	DNA-directed RNA p
64	117	5.8	785	2 T09451	hemagglutinin, pha
65	116.5	5.8	338	1 TVMSEB	transforming prote
66	116.5	5.8	859	2 H70327	DNA mismatch repai
67	116.5	5.8	1790	2 S67593	transport protein
68	116	5.8	393	2 P00479	pistil extensin-li
69	116	5.8	451	2 B70792	hypothetical prote
70	116	5.8	462	1 LUD07	annexin VII - slim
71	116	5.8	504	2 A49467	occludin - chicken
72	116	5.8	513	2 T03916	hypothetical prote
73	116	5.8	609	2 S46019	YSM1 protein - yea
74	116	5.8	821	2 S67087	hypothetical prote
75	115.5	5.8	177	2 T47549	hypothetical prote
76	115.5	5.8	428	2 T24769	hypothetical prote
77	115.5	5.8	1189	2 T42726	guanine nucleotide
78	115	5.7	240	2 D70894	probable pra prote
79	114.5	5.7	240	2 A24264	proline-rich prote
80	114.5	5.7	272	2 T25608	hypothetical prote
81	114.5	5.7	324	2 S13497	cAMP-binding prote
82	114.5	5.7	398	2 T34947	hypothetical prote
83	114.5	5.7	554	2 T86244	hypothetical prote
84	114.5	5.7	817	2 S51342	hypothetical prote
85	114.5	5.7	1633	2 JC5056	verprolin - yeast
86	114.5	5.7	1940	1 S04090	polybromo 1 - chic
87	114	5.7	342	2 A24263	myosin heavy chain
88	114	5.7	1017	2 T31354	myosin heavy chain
89	114	5.7	1179	2 F71190	probable potassium
90	114	5.7	1464	1 CGHU1S	collagen alpha 1(I
91	113.5	5.7	347	1 TVHUJUS	transforming prote
92	113.5	5.7	1046	2 T42734	cytoplasmic linker
93	113.5	5.7	1487	1 CGHU6C	collagen alpha 1(I
94	113	5.6	242	2 S35060	tropomyosin - hydr
95	113	5.6	481	2 F86208	protein F22G5.30 [
96	113	5.6	1638	2 A42091	transcription acti
97	112.5	5.6	199	2 S14981	extensin class I (
98	112.5	5.6	324	2 G86222	hypothetical prote
99	112.5	5.6	358	2 C42026	cyclic AMP respons
100	112.5	5.6	388	2 JC5437	spliceosome-associ

ALIGNMENTS

RESULT 1
H88650
protein C09G12.9 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: H88650
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999
A:Accession: H88650
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <STO>
A:Cross-references: GB:chr_IV; PIDN:AAC25822.1; PID:G3294495; GSPDB:GNO0022; CESP:C09G12.9
C:Genetics:
A:Gene: C09G12.9
A:Map position: 4

Query Match 32.9%; Score 658; DB 2; Length 404;
Best Local Similarity 36.7%; Pred. No. 5.3e-31;
Matches 151; Conservative 74; Mismatches 121; Indels 66; Gaps 11;
QY 6 KYRDLTVETNVITLYKDLKPVLDYVFNDSRELMLNLTGTIPVYRNTYNIPICLW 65
DB 16 KYADSAKKDIIIGALSQFKDLSPGTDTFMFPDQKRRRTAFRLKGTIPVYKGCYNIPVTY 75
QY 66 LDTYPNPPICPVKPTSTMTIKGHVDANGKIYLPYLHEWKHPQSDLLGLLOVLMVFP 125
DB 76 LWDTHPYAPICVYNTSTW---ESEHVNKEGVFLPYLNEWRFPFGYDLISGLLQM----- 127
QY 126 GDEPPVFSRPIASYPYQATGP-----PNTSYMPCMGPGGISPYPSG-----YP 169
DB 128 -----IFARSAANSATNATNSFNSAGSSASTTTPYSSQPTWP---TPYTSAGRAPY 179
QY 170 PNPSGYPGC-----PYPPG-----GPYPATTSQY---PSQPPV---TTVG 204
DB 180 PSSTPYPSAGMGVYPMVNVQSTFYDMGASGSPYPSAGSNAPPYPPRPPVTAQTSVS 239
QY 205 PSRDGTISEDIRASLISAVSKLRWKEMKEMDRAQAEINAKLKTEEDLKKGHQKLEEMV 264
DB 240 SSSGGTQADTIRASVMSAVEEIKIRAKLRERMGNTSAEMASIRTSDELREGQOKLRML 299
QY 265 TRLDQAEVNDKNIELKKDEELSSALEKMNQSENNDIDEVITPTAPLYKQILNLAYE 324
DB 300 BELETQRSSIQTACEIYTAKAEALXALSD-AGTDAPPIDEALDAAPFLHRQIVLNYAK 358
QY 325 ENAIEDTIFYLGEARRGVLDLDFVLKXVRLLSKQFQLRALMQKARTAGL 376
DB 359 DLTC-----QSLKKRQITLAELYHVRDVSREQFIYRATMGKCRRTAGL 402

RESULT 2
I48283
gene CC2 protein - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C:Accession: I48283
R:Maucuer, A.; Camonis, J.H.; Sobel, A.
Proc. Natl. Acad. Sci. U.S.A. 92, 3100-3104, 1995
A:Title: Statmin interaction with a putative kinase and coiled-coil-forming protein domain
A:Reference number: I48282; PMID:95241452; PMID:7724523
A:Accession: I48283
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-83 <RES>
A:Cross-references: EMBL:X82319; NID:G791075; PIDN:CAAS7762.1; PID:G791076
C:Genetics:

A:Gene: CC2

Query Match 19.4%; Score 388; DB 2; Length 83;
Best Local Similarity 97.6%; Pred. No. 3.2e-16;
Matches 80; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 231 RMKEEMDRAQAEINAKLKTEEDLKKGHQKLEEMVTRLDQAEVNDKNIELKKDEELSS 290
DB 1 RMKEEMDRAQAEINAKLKTEEDLKKGHQKLEEMVTRLDQAEVNDKNIELKKDEELSS 60

QY 291 ALEKMNQSENNDIDEVIPTA 312
DB 61 ALEKMNQSENNDIDEVIPTA 82

RESULT 3
T34520
hypothetical protein DKFp564J157.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34520
R:Boecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A:Reference number: Z21539
A:Accession: T34520
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-169 <BLO>
A:Cross-references: EMBL:AL122042
A:Experimental source: fetal brain; clone DKFp564J157
C:Genetics:
A:Note: DKFp564J157.1

Query Match 7.7%; Score 153.5; DB 2; Length 169;
Best Local Similarity 30.4%; Pred. No. 0.027;
Matches 42; Conservative 18; Mismatches 45; Indels 33; Gaps 6;
QY 126 GDEPPVFSRPIASYPYQATGPNTSYMPCMGPGGISPYSGYPNP---SGYPGC----- 178
DB 43 GGSNPAHPPIPNFPFGPCPP-----PGAPGNPAPSPGPPHPVPGYPCQPLG 96
QY 179 PYPGPGYPATTSQVPSQPPVTVVPSRDGTISEDIRASLISAVSKLRWKEMKEMDR 238
DB 97 FYPPPPYPPPA-----PGIPPNPLAPGMVGP-----AVIVDK---KMQKCKKK 136
QY 239 AQAEINAKLKTEEDLKKG 256
DB 137 AHKMKHKHQKHXYHKG 154

RESULT 4
LHU7
annexin VII, long form - human
N:Alternate names: synexin
N:Contains: annexin VII, long form; annexin VII, short form
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1992 #sequence_revision 26-Jan-1996 #text_change 22-Jun-1999
R:Shirvan, A.; Srivastava, M.; Wang, M.G.; Cultraro, C.; Magendzo, K.; McBride, O.W.; P.
Biochemistry 33, 6888-6901, 1994
A:Title: Divergent structure of the human synexin (annexin VII) gene and assignment to
A:Reference number: A54467; MUID:94264005; PMID:7515686
A:Accession: A54467
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-488 <SHI>
R:Burns, A.L.; Magendzo, K.; Shirvan, A.; Srivastava, M.; Rojas, E.; Alijani, M.R.; Pol
Proc. Natl. Acad. Sci. U.S.A. 86, 3799-3802, 1989
A:Title: Calcium channel activity of purified human synexin and structure of the human
A:Reference number: A32554; MUID:89264510; PMID:2542947
A:Accession: A32554
A:Molecule type: mRNA
A:Residues: 1-145,168-488 <BUR>

submitted to the Protein Sequence Database, September 1996

A;Reference number: S74288
A;Accession: S74288
A;Molecule type: DNA
A;Residues: 1-212 <WED>
A;Cross-references: EMBL:X59720; NID:gl907116; PIDN:CAA42351.1; PID:e309039; PID:gl90713f
R;Oliver, S.G.; Anwar, R.; Brown, A.; Gent, M.E.; Indge, K.J.; James, C.M.; Stateva, L.I
submitted to the Protein Sequence Database, March 1992

A;Reference number: S19337
A;Accession: S19410
A;Molecule type: DNA
A;Residues: 1-90, 'GEYS', 95, 'TA', 98-99, 'AQVSTP', 106, 'TAIASTTTSTGI' <OLI>
A;Cross-references: EMBL:X59720; MIPS:YCL008c
A;Note: this sequence has been revised in reference S74288

C;Genetics:
A;Gene: SGD:STP22
A;Cross-references: SGD:S0000514
A;Map position: 3L
A;Note: YCL008c

Query Match 7.3%; Score 145.5; DB 2; Length 212;
Best Local Similarity 24.5%; Pred. No. 0.1; Mismatches 38; Indels 41; Gaps 11;
Matches 56; Conservative 38; Mismatches 100; Indels 41; Gaps 11;

Qy 64 LWLLDTYPYNPPIC-----EVKPTSSMTIKGKHVDANGKIYLPYLHEWKHPQSDLLGL 117
Db 1 MWVSPYVVKPPFFISINLENFDMNTISSSLPIQEIYDSNGWIALPILHCWDPAAMNLIW 60

Qy 118 IQVMIVVFGDEPPVFSRISASYPYQATGPNTSYMPGMPGIGSPYSGYPNPSGYPG 177
Db 61 VQELMSLL-HEPPQDQAPSLPFPKPTQVQEQNTPLPFPKPS----PHLKPPLP----- 110

Qy 178 CPVPPGGPYPATTSQSPQSPVPTVGPSPD-----GTISEDITRASLISAVSDKLKW 230
Db 111 ---PPPPQFASNALDMDMN-TDSTNNHEMMQNTQVNVNELYRED-VYVADKIIT 165

Qy 231 R---MKEMDRAQALNALKRTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKK 284
Db 166 RQTVMQESIARFH-EIITA-----IDKNHLRAVEQA--IEQTMHSLNAQIDVLNRK 212

RESULT 6
A53152
annexin XI - human
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
C;Accession: A53152
R;Misaki, Y.; Pruijn, G.J.M.; van der Kemp, A.W.C.M.; van Venrooij, W.J.
J. Biol. Chem. 269, 4240-4246, 1994
A;Title: The 56K autoantigen is identical to human annexin XI.
A;Reference number: A53152; MUID:94140847; PMID:7508441
A;Accession: A53152
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-505 <MIS>
A;Cross-references: GB:L19605; NID:g457128; PIDN:AAA19734.1; PID:g457129
C;Genetics:
A;Gene: GDB:ANX11
A;Cross-references: GDB:313076
A;Map position: 9q11-9q22
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid bin
F;203-274/Domain: annexin repeat homology <AX1>
F;275-346/Domain: annexin repeat homology <AX2>
F;358-430/Domain: annexin repeat homology <AX3>
F;434-505/Domain: annexin repeat homology <AX4>

Query Match 7.1%; Score 142; DB 2; Length 505;
Best Local Similarity 28.0%; Pred. No. 0.47;
Matches 63; Conservative 19; Mismatches 91; Indels 52; Gaps 10;
Qy 111 QSDLLGLIQVMIVVFG--DEPPVFSRPTISASYPYQ--ATGPPNTSYMPGMPGIGSPYS 166

Db 51 QDYLSCMAANSGTFFGANNPNLPGAPGAGYPPVPPGGQPPSAQQQVPPYGMVPPPG 110
QY 167 GYPPN-PS-GYPGCPYP-----PGGPYPATTSSQYPSQPPVTTVG-----PSR 207
Db 111 GNPPSRMPSYPYPGAPVPGQPMPPPGQPPGAPGQPPVTPGQPPVPLPQQQVPSY 170
QY 208 DGTISDDTI-----RASLISAV-----SDKLWRMKEEMDRQAELNAKRT 249
Db 171 PGYPGSGTVPAPPTQFGSRGTHITDAPGDPDLRAEVLKAMKGGTDEQALIDCLGR 230
QY 250 EBDLKGKHGLEMVRTLDQEAQVAEVDKNIELLKKDEELSSALEK 294
Db 231 SN--KORQQLLSFKTAYGK-----DLIKLAKSELSGNFEK 264

RESULT 7
S29170
annexin VII - mouse
N/Alternate names: synexin
C/Species: Mus musculus (house mouse)
C/Date: 25-Feb-1994 #sequence revision 01-Sep-1995 #text_change 13-Aug-1999
C/Accession: S29170; S46209; S51173
R/Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
Biochem. J. 289, 735-741, 1993
A/Title: Mouse synexin (annexin VII) polymorphisms and a phylogenetic comparison with other species
A/Reference number: S29170; PMID:93168121; PMID:7916616
A/Accession: S29170
A/Molecule type: mRNA
A/Residues: 1-463 <ZHA>
A/Cross-references: EMBL:L13129
R/Zhang-Keck, Z.Y.; Srivastava, M.; Kozak, C.A.; Caohuy, H.; Shirvan, A.; Burns, A.L.; H. Biochem. J. 301, 835-845, 1994
A/Title: Genomic organization and chromosomal localization of the mouse synexin gene.
A/Reference number: S46209; PMID:94330961; PMID:80533909
A/Accession: S46209
A/Molecule type: DNA
A/Residues: 1-463 <ZHF>
R/Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
submitted to the EMBL Data Library, May 1993
A/Reference number: S51173
A/Accession: S51173
A/Molecule type: mRNA
A/Residues: 1-144,'S',146-303,'A',305-463 <ZHW>
A/Cross-references: EMBL:L13129; NID:G293293; PIDN:AAA37238.1; PID:G293294
C/Genetics:
A/Gene: MGI:Anx7
A/Cross-references: MGI:88031
A/Map position: 14
A/Introns: 13/3; 87/1; 121/1; 142/3; 177/1; 208/3; 246/3; 303/3; 360/3; 386/1; 423/3
C/Suprafamily: annexin VII; annexin repeat homology
F/163-234/Domain: annexin repeat homology <AX1>
F/235-306/Domain: annexin repeat homology <AX2>
F/318-390/Domain: annexin repeat homology <AX3>
F/394-463/Domain: annexin repeat homology <AX4>

Query Match 7.0%; Score 141; DB 2; Length 463;
Best Local Similarity 22.0%; Pred. No. 0.49;
Matches 76; Conservative 43; Mismatches 90; Indels 136; Gaps 17;
QY 139 SYPPYQATGPPNTSYMP-----GMPGGISP-----YPSGY-----PPNPSGYPGCPY 180
Db 45 AYPPAPSGGYGAGGYPAPGGYPAGGYGALSPGGPPAYPGGQFGGAPPGGAGFGSGYFQ 104
QY 181 PP-----GGP-----YPATSSQYPSQPPVTTVGPSRDGTI-----SEP 214
Db 105 PPAQSYGGGPAQVPPVPGGFGGMPQSYPG-GGAPYPSQPAAMTQG--TQGTILPASNFD 161
QY 215 TIRASLI-----SAVSDKLWRMKEEMDRQAELNAL--KRTEDLKK-----255
Db 162 AMRDAELRKMKGFGTDEQALVDVSVNSRNDQRQIKAAFKTMYGKDLIKLKSLSGN 221
QY 256 -----GHQK-----LEEMVTRLDOEVAEVDKNIELLKKKD 285

Db 222 MEELILALFMPSTYYDAWSLRKAMQAGTQERVLIELCTRTNQIRDIRVRCVQLEFGRD 281
QY 286 -----EELSSALEKM-----ENQSENNDDEVILPTAPLYKQILNLYAEENALE 329
Db 282 LEKDIRSDTSGHFERLLVSMCGNDRDQSVNHQ-----AQEDA-- 321
QY 330 DTIFVLGEALRRGVLDLDFVLKHLVRLSRKQFQLRALMOKARKTA 374
Db 322 QRLYQAGE---GRLGTDESCFNMLLATRSFPQLKATMEAYSERMA 362

RESULT 8
T39608
zinc finger transcription factor - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text_change 02-Sep-2000
C/Accession: T39608
R/Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A/Reference number: Z21866
A/Accession: T39608
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-827 <LYN>
A/Cross-references: EMBL:AL023554; PIDN:CAA19035.1; GSPDB:GN00067; SPDB:SPBC16G5.16
A/Experimental source: strain 972h; cosmid c16G5
C/Genetics:
A/Gene: SPDB:SPBC16G5.16
A/Map position: 2
A/Introns: 30/3; 41/1
C/Suprafamily: GAL4 zinc binuclear cluster homology
F/11-47/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 6.9%; Score 138.5; DB 2; Length 827;
Best Local Similarity 25.2%; Pred. No. 1.4;
Matches 60; Conservative 28; Mismatches 99; Indels 51; Gaps 11;
QY 13 RETVNVITLYKDLKPVLDYSVFNDGSSREL---MNLGTIPVPYRGNTYNIP-----ICL 64
Db 589 RESTYYSTIMEAKNLLIAYDMNSGTENLDATPDVTGQLPNNFSORTSNIPREFPQAI 648
QY 65 WLDDT-YP--YNP-----PICFVKPTSSMTIKTKHVDANGKIYLYLHMKHPOSDDL 115
Db 649 FYSADAPGYNPAQFQNAFTNPMPTYGRTQDQSYPRQNG--YPSYSDGNVYHDRV 706
QY 116 GLIQWIVVVGDEPPVFSR-----PISASYPYQATGPPNTSYMPGPGISP 163
Db 707 -----INYGSMFTFANGFYVNTVSPVFFYNTSYPPYMSD-----TSNMPQAFAYSQ 755
QY 164 YPSGYPPNPSGYPGCPYPPGG---PYPATSSQYPSQPPVTTVGPSRDGTISEDTIR 217
Db 756 YPQHPFPPLSQMLPLPTSGVMVMAFGAAGKSGMPYFIQF-----PSMINQVAYPIVR 808

RESULT 9
LURB11
annexin XI - rabbit
N/Alternate names: calyculin-associated annexin protein CAP-50
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 16-Jun-2000
C/Accession: JH0694; PH0950; A38250; PS0263
R/Tokumitsu, H.; Mizutani, A.; Muramatsu, M.; Yokota, T.; Arai, K.; Hidaka, H.
Biochem. Biophys. Res. Commun. 186, 1227-1235, 1992
A/Title: Molecular cloning of rabbit CAP-50, a calyculin-associated annexin protein.
A/Reference number: JH0694; PMID:92378579; PMID:1380798
A/Accession: JH0694
A/Molecule type: mRNA
A/Residues: 1-503 <FOK>
A/Cross-references: DBJ:DI0883; NID:G471147; PIDN:BAA01705.1; PID:G471148
A/Experimental source: lung
A/Accession: PH0950
A/Molecule type: protein

A;Residues: 104-141;213-231;254-262;270-280;285-309;319-337;429-448;478-492 <TK2>
R;Tokumitsu, H.; Mizutani, A.; Minami, H.; Kobayashi, R.; Hidaka, H.
J. Biol. Chem. 267, 8919-8924, 1992
A;Title: A calyculin-associated protein is a newly identified member of the Ca(2+)/phosphatase family
A;Reference number: A38250; MUID:92250478; PMID:1533622
A;Accession: A38250
A;Molecule type: protein
A;Residues: 104-141;213-231;254-262;270-280;285-309;319-337;429-448;478-492
C;Comment: This protein binds specifically to calyculin in a Ca2+ dependent manner.
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid binding
F;201-272/Domain: annexin repeat homology <AX1>
F;212-228/Region: endonexin fold #status predicted
F;273-344/Domain: annexin repeat homology <AX2>
F;284-300/Region: endonexin fold #status predicted
F;356-428/Domain: annexin repeat homology <AX3>
F;368-384/Region: endonexin fold #status predicted
F;432-503/Domain: annexin repeat homology <AX4>
F;443-459/Region: endonexin fold #status predicted
F;58/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 6.8%; Score 136.5; DB 1; Length 503;
Best Local Similarity 24.1%; Pred. No. 0.98;
Matches 81; Conservative 39; Mismatches 117; Indels 99; Gaps 18;
QY 126 GDEPPV-----PSRPISASYPYQATGPPNTSYMPGPGISYPSG---YPPNPSG-YPG 177
Db 79 GGYPPVPPGGFGQP-----PPTQSPVPPGVYPP--PGG-NP-PSGVSPYPPFGAPVPG 129
QY 178 CPYP-----PGGYPATTSQYSPSPVPTVG-----PSRDGPTISBDTI-----R 217
Db 130 QEMPPPHQPPGYPGQLPVTPYGGSPVPPGQPMPSYPGYVPGSGTVPAPVPPVQFGNR 189
QY 218 ASLISAV-----SDKLWRKMEMDRAQAEINL-----KTEEDLAK-----CHQKL 260
Db 190 GTITDASGFDPLRDAEVLKAMKGFCTDEQAIDICLSRKNQKQKQILLSFTAYGKULI 249
QY 261 ESMVTELD-----QVAVEDKN-----IELLKKKDEELSSALE 293
Db 250 KOLKSELNSNFEKTIILAMKTPILFYAEIKAEKAGTDEACLIILASRNSHEIRLN 309
QY 294 KVENQSENNDIDRVI--IPTAPLYKQILNLYAENAEIETD-----IFVYLGAL 339
Db 310 KAYKTEFKTLBAIRSDTSGHFQRLILSLSQNRDESTNVDMSLVQRDVQELYAAGE-- 367
QY 340 RRGVIDDLVFLKHVRLLSRKQFQALRMQKARTAG 375
Db 368 --NRLGTDSEKFNALVLCRSRAHLVAVFNEYQRMGTG 401
RESULT 10
T14192
extensin homolog T28D5.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-Jan-2000
A;Accession: T14192
R;Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft, J.; et al. 1999
submitted to the Protein Sequence Database, August 1999
A;Reference number: 217931
A;Accession: T14192
A;Molecule type: DNA
A;Residues: 1-437 <BEV>
A;Cross-references: EMBL:AL109819
A;Experimental source: cultivar Columbia; BAC clone T28D5
C;Genetics:
A;Gene: ATSP:T28D5.70
A;Map position: 4
C;Superfamily: hydroxyproline-rich glycoprotein
Query Match 6.8%; Score 135.5; DB 2; Length 437;
Best Local Similarity 26.8%; Pred. No. 0.94;
Matches 49; Conservative 16; Mismatches 75; Indels 43; Gaps 8;

QY 50 PVPYRGNTYNIPICLWLLDTYV-----NPPICFVKPTSSMTIKTGKHDANGKIYL----- 101
Db 43 PPFY---TSPFPSPYVYKSPYVSSPPYVAYSPSPYVYKSPYVSSPPYVAYSP 99
QY 102 --PYLHEWKHPQSDLLGLIOMIVVFGDEPPVFSRPIASYSY-----PPYQATGPPNTSYMP 155
Db 100 PPFV--YKSP-----YVYSSPPYVAYSPSPYVYKSPYVSSPPYVYSS 146
QY 156 GMPGGISYPSGY--PNNPSGYGCPY-----PPGGYPATTSQYSPQPPVTT 202
Db 147 PPFVAYSPPPYVAYSPSPYVYKSPYVSSPPYVAYSPSPYVYKSPYVSSPPY 206
QY 203 VGP 205
Db 207 YSP 209
RESULT 11
D70509
hypothetical protein rv1233c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: D70509
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: D70509
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-198 <COL>
A;Cross-references: GB:Z98260; GB:AL123456; NID:G3261826; PIDN:CAB10930.1; PLE:el299942;
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: RV1233c

Query Match 6.7%; Score 134; DB 2; Length 198;
Best Local Similarity 32.8%; Pred. No. 0.44;
Matches 42; Conservative 12; Mismatches 42; Indels 32; Gaps 7;
QY 126 GDEPPVFSRPI-----ASYPP--YQATGPPNTSYMPGPGISP----- 163
Db 17 GGGPPVGERPPEQPIADAPWAPPASSPMANHPHPATPPSGYP-PAYQPGYPTGYPPMP 75
QY 164 ---YPSGYPP---NPSGYGCPYPP--GGYPATTSQYSPQPPVTTVTPSRDGTISBDTI 216
Db 76 GGYAPFGYPPPGTSSAGYDIPYPPMPVPGSGPGYVPEPGYLDGYGSPQGNMTALV 135
QY 217 RASLISAV 224
Db 136 --SLISAL 141

RESULT 12
T37550
hypothetical coiled-coil protein [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 08-Sep-2000
C;Accession: T37550
R;Connor, R.; Churcher, C.M.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21723
A;Accession: T37550
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-485 <CON>
A;Cross-references: EMBL:AL132714; PIDN:CAB59797.1; GSPDB:GN00066; SPDB:SPAC11H11.01
A;Experimental source: strain 972h-; cosmid c11H1
C;Genetics:
A;Gene: SPDB:SPAC11H11.01

380 VLDRDVAQLVAEAEV ---EPALPAAEARWDN 408

 \mathbf{Y}

Y 168 --YPPNPSGYPGCPYPGGP-----YPATTSSQVPSQP----- 198

```

Db 300 KLIPPQPTGRS-PAPPAVPPAASVPMPEPQTQSPQPAQPMVPLHOKQSRITPIQKPR 358
QY 199 -----PVTTPGSRDGTISE----- 213
Db 359 GLDPVEILOREYRLQARIAHRIQELNMLPGSLVGDULTKTKVITIELKDLRLNFORQLRQE 418
QY 214 -----DTI-----RASLISA-VSKLWRMKEEMDRA-----QAEINLAL 246
Db 419 VVVCMRDRTALETALNVKAYKRKQSLREARITKLEKQKQIEQERKRRKHQOYLNSI 478
QY 247 KRTEEDLKKGHOKLEEMVTRLDQAEVVDKNIELKKDBELSSALEMENQSENNDIDE 306
Db 479 LQAKDFEYHRSVTGTITQKLTAKVSTYHANTEREQKENE-----RIEKERIRR 528
QY 307 VIIPTAPLYKQILNLYAEENAIEDTIFVLGALRRGVIDL-DVFLKHVRLLSRKQFOLRA 365
Db 529 LMADEEGYRLIDQKDKR-----LAYLLOQTDYVANLTELVRQHKASQVAKEIKKKK 583
QY 366 LMQKARKTAG 375
Db 584 TKKKAENAE 593

RESULT 16
T18531
C:Species: Hirudo medicinalis (medicinal leech)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C:Accession: T18531
J: Huang, Y.; Jellies, J.; Johansen, K.M.; Johansen, J.
A: Title: Differential glycosylation of Tractin and LeechCAM, two novel Ig-superfamily me
A: Reference number: Z18951; MUID: 97362067; PMID: 9214388
A: Accession: T18531
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-1880 <HUA>
A: Cross-references: EMBL:U92813; NID: g2275259; PID: g2275260; PIDN: AAC47654.1

Query Match 6.5%; Score 131; DB 2; Length 1880;
Best Local Similarity 30.4%; Pred. No. 10;
Matches 48; Conservative 13; Mismatches 71; Indels 26; Gaps 7;

QY 142 PYQATGP-----PNTSYMPGMPGGISPYPSGYPNPSPGYPGCPYPPGPGY-----PAT 189
Db 1596 PYGPGGPGYPGPGYGGPGGG--PYGPGKPGGG-PGPGYGGPGYGGPGHGGPGS 1652
QY 190 TSSQVPSQPFVTVGSRDGTISEDTIRASLISAVSDKLKW--RMKEEMDRAQAEINLAL 247
Db 1653 PRDLPDSTSPKTDKGIKIRG--KDKPKSRKRGQRDKGKWTILFEE-----EKNWTIN 1703
QY 248 RTEEDLKKGHOKLEEMVTELDOEVAEVDKNIELKKKD 285
Db 1704 LTNLDGSTYPRVATNNLNGKEIKSEPKRMVITAKKE 1741

RESULT 17
I53043
transforming protein fos-B - human
N: Alternate names: G0S3
C: Species: Homo sapiens (man)
C: Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 16-Jul-1999
C: Accession: I53043
R: Siderovski, D.P.; Blum, S.; Forsdyke, R.E.; Forsdyke, D.R.
DNA Cell Biol. 9, 579-587, 1990
A: Title: A set of human putative lymphocyte G0/G1 switch genes includes genes homologous
A: Reference number: I53043; MUID: 91103878; PMID: 1702972
A: Accession: I53043
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-338 <RES>
A: Cross-references: GB:I49169; NID: g1082037; PIDN: AAB53946.1; PID: g1082038
```

```

C: Genetics:
A: Gene: G0S3
C: Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
F: 150-190/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 6.5%; Score 130.5; DB 2; Length 338;
Best Local Similarity 28.3%; Pred. No. 1.3;
Matches 51; Conservative 18; Mismatches 68; Indels 43; Gaps 7;

QY 113 DLLGLIQVMIV-----VFGDEPPVFSRPIASVPPYQATGPPNTSY-MPGMPPGGIS 162
Db 66 DLQMLVQPTLISSMAQSQGQPLASQPPVVD-----PYDM--PGTSYSTPGMSGYSS 114
QY 163 PYPGYPNPSPGYPGCPYPGYPATTSYQSPQPPVTVTVGSRDGTISEDTIRASLIS 222
Db 115 GGASG-----SGGPSTSGTSGPGPARPARPRPREETILTPEE---EKRRVRERNK 166
QY 223 AVSKLWRMKEEMDRAQAEINLALKRTEDLKKGHQKLEEMVTRLDQAEVVDKNIELK 282
Db 167 LAAAKCRNRRLDRLQAEYD-----QLEBEKAELESEIAELQEKERLE 212

RESULT 18
S45252
SNF2beta protein - human
C: Species: Homo sapiens (man)
C: Date: 10-Dec-1994 #sequence_revision 17-Nov-1995 #text_change 02-Aug-2002
C: Accession: S45252
R: Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
Nucleic Acids Res. 22, 1815-1820, 1994
A: Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brahm
A: Reference number: S45251; MUID: 94268902; PMID: 8208605
A: Accession: S45252
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-1647 <CHI>
A: Cross-references: GB:D26156; NID: g505087; PIDN: BAA05143.1; PID: g505088
C: Superfamily: human SNF2alpha protein; bromodomain homology
F: 1485-1540/Domain: bromodomain homology <BRO>

Query Match 6.5%; Score 130; DB 2; Length 1647;
Best Local Similarity 21.7%; Pred. No. 9.8;
Matches 61; Conservative 27; Mismatches 71; Indels 122; Gaps 11;

QY 126 GDEPPVFSRPIASVPPYQATGPPNTSYMPGMP-----GGISPYPSG----- 167
Db 243 GPAPNYSRPHGMGPNPMPGP--SGVPPGPGQPPGPGPKPWPPEGPMANAAFTSPQ 300
QY 168 --YPPNPSPGYPGCPYPPGPG--YPAATTSQVPSQPP-----VTVGSPSR 207
Db 301 KLIPPQPTGRS-PAPPAVPPAASVPMPEPQTQSPQPAQPMVPLHOKQSRITPIQKPR 359
QY 208 -----DGTISED-----TI----- 216
Db 360 GLDPVEILOREYRLQARIAHRIQELNMLPGSLAGDLRTKATIELKALRLNFORQLRQE 419
QY 217 -----RASLISA-VSKLWRMKEEMDRA-----QAEINLAL 246
Db 420 VVVCMRDRTALETALNVKAYKRKQSLREARITKLEKQKQIEQERKRRKHQOYLNSI 479
QY 247 KRTEEDLKKGHOKLEEMVTRLDQAEVVDKNIELKKKDEE 287
Db 480 LQAKDFEYHRSVTGTITQKLTAKVATYHANTEREQKENE 520

RESULT 19
T42731
atrophin-1 related protein - rat
C: Species: Rattus norvegicus (Norway rat)
C: Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C: Accession: T42731
R: Khan, F.A.; Margolis, R.L.; Loev, S.L.; Sharp, A.H.; Li, S.H.; Ross, C.A.
submitted to the EMBL Data Library, December 1995
```

A;Description: CDNA sequence and expression of an atrophin-1 (DRPLA disease gene) related
A;Reference number: Z22250

A;Accession: T42731

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1006 <KHA>

A;Cross-references: EMBL:U44091; NID:g1297310; PID:g1209103; PIDN:AAA99970.1

C;Genetics:

A;Gene: ARP

Query Match 6.4%; Score 128.5; DB 2; Length 1006;
Best Local Similarity 22.1%; Pred. No. 6.6;
Matches 69; Conservative 42; Mismatches 112; Indels 89; Gaps 16;

QY 44 NLCTTIVPVVRGNTYNIPICLWLLDTYPYNNPICFVKPTSMITKTGHVDANGKIYLPY 103

DB 388 HLSC--ESPFSMNA-NLP-----PPPA-LKPLSLSTH-----HPPS 420

QY 104 LHE---WKHPQSDLL-----GLIQMIVVFGDEPPVFSRPTISASYPYPYQATGP-PNT 151

DB 421 AHPPPLQLMPQSOPLPSSPAQPGCLTQSGL-----PPPAASHPTTGLHQVPSQSPFPQH 476

QY 152 SYNPGMGGISPPSPGYP-----PNPSPVPGC-----PYPPGSPYPAT----- 189

DB 477 FVFGGPPPTIP-PSCEPTSTPPAGPSSSQPPPCSAAVSSGNNVPGAPSCPLPAVQIKEE 535

QY 190 --TSOYPSQPPVTVGSPSDGTISEDTIRASLISAVSDKLRWKMKEMDR-----AQAE 242

DB 536 ALDEAEPEFPFPPSPSPPTVVDTPSHASQSA-----RFYKHLDRGYNNSCARTD 587

QY 243 L-----NALKREEDLKGKHKLEEMVTRLDQEAEDVDKNIELLKKKDEELSSALEK 294

DB 588 LYFEMPLAGSLAKKREAEIAKAKAEQAKEREEREKEKEKERERERERERAEARAQK 647

QY 295 MENQSENNDIDE 306

DB 648 ASSSAHEGRSLSD 659

RESULT 20

H84809

hypothetical protein At2g38830 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: H84809

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosomes 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: H84809

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-139 <STO>

A;Cross-references: GB:AE002093; NID:g4585902; PIDN:AAD25564.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g38830

A;Map position: 2

Query Match 6.4%; Score 128; DB 2; Length 139;

Best Local Similarity 24.1%; Pred. No. 0.64;

Matches 33; Conservative 41; Mismatches 49; Indels 14; Gaps 2;

QY 236 MDRAQAEALNALKTEEDLKGKHKLEEMVTRLDQEAEDVDKNIELLKKKDEELSSALEKM 295

DB 1 MERSEEEIENLKLQEVQRSEVAKSIITEIERGTUKVRALKUKEDSDVLTTWVWN 60

QY 296 ENQSENNDIDEVIPTAPLYKQILNYAE-----ENAEITFIYLGAEALRRGVLDLDF 349

DB 61 YLKLTSMDMGRI-----EEMFEISEVEGLAGDAIEDVLRVLVEAEERGELEIGSY 112

QY 350 LKHVLLSKKQQLRAL 366

DB 113 LKQVRVLAREQOFFLKLHL 129

RESULT 21

T42730

Bassoon protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C;Accession: T42730

R;Dieck, S.; Sammarti-Vila, L.; Langhaese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex

J. Cell Biol. 142, 499-509, 1998

A;Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively localized

A;Reference number: Z22249; MUID:98345363; PMID:9679147

A;Accession: T42730

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-3942 <DIE>

A;Cross-references: EMBL:Y17034; NID:g3413809; PIDN:CAA76598.1; PID:g3413810

A;Experimental source: strain 129 SVJ

C;Genetics:

A;Map position: 9F1

A;Introns: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1

A;Note: bassoon

C;Function:

A;Description: may be involved in cytomatrix organization at the site of neurotransmitter

A;Note: component of the presynaptic cytoskeleton

C;Keywords: coiled coil; zinc finger

Query Match 6.4%; Score 128; DB 2; Length 3942;

Best Local Similarity 21.4%; Pred. No. 37;

Matches 72; Conservative 54; Mismatches 117; Indels 94; Gaps 16;

QY 40 RELMNTGTIPVP-----YRGNTYNIPICLWLLDTYPYNNPICFVKPTSMITKTGHVDA 95

DB 2183 RQLPSTATVRAADGMIV--STINTPLAATLPIT---TQPASVLRPM-----V 2225

QY 96 NGKIYLPYLHEWKHPQSDLLGLIQMIVVFGDEPPVFSR-----PISAS 139

DB 2226 RGMVREPVSG-----GVTAVPLTSLTRVPMIAPRVPLGPAFLYRYPAPRFFPIASS 2276

QY 140 YPPYQ-----ATGPPNNTSYMPGMPGGIS---PYPSGVPPNPSPGPGCPYP 181

DB 2277 VPAEGFVVLGKPAATKASGAGGPPR---PELPAGVAREEPSTTAPAVIKAPVAPAP 2332

QY 182 PGGPYATTSQVPSQPPVTVGSPSDGTISEDTIRASLISAVSDKLRWKMKEE---MD 237

DB 2333 --GPAPAPPPGOKPAGAAAGSG---SGVLSRP---ASEKEEASQEDRQKQEQQLLQLE 2384

QY 238 RAQAELNALK--RTEEDLKGKHKLEE-----VYTRLDQEAEDVDKNIELLKKKDEEL 288

DB 2385 RERVELEKRLQLQLEELERVELQHRHEEQQLVQREIQLELTQIKQHVLOQQQBERQA 2444

QY 289 SSALKEKMNQSENNDIDEVIPTAPLYKQILNYAE 325

DB 2445 QFALQREQLAQQRLQLEQI-----QQLOQQQLQLEEE 2476

RESULT 22

G01763

atrophin-1 - human

C;Species: Homo sapiens (man)

C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998

C;Accession: G01763

R;Margolis, R.L.

submitted to the EMBL Data Library, March 1995

A;Reference number: G08343

A;Accession: G01763

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1184 <MAR>

A;Cross-references: EMBL:U23851; NID:g915325; PID:g915326

C;Genetics:

Db 258 GDMSDIELEASKMKEOADEMRSLAEFVFNERRAKEEELQGLEIEVEEQKLN EAVTHAMD 317

A:Gene: GDB:DRPLA; B37
A:Cross-references: GDB:270336; OMIM:125370
A:Map position: 12p24

Query Match 6.3%; Score 127; DB 2; Length 1184;
Best Local Similarity 23.6%; Pred. No. 9.8;
Matches 64; Conservative 32; Mismatches 93; Indels 82; Gaps 13;

QY 70 YPNPPICFVKTSMTKIT-----GKHVDANGKI----- 99
|||:|||:|||:|||:
Db 593 YPF-PPVPTV-TTSSATLSTVIATVASSPAGYKXTASPGPPPYGKRAPSOGAYKTATDPG 650
|||:|||:|||:|||:
QY 100 YLPYLHEWKHPQSDLLGLIQVMWIFGDEPPVFESRPISASYPYPYQATGPP--NTSYPMGM 157
|||:|||:|||:|||:
Db 651 YKP-----GSPFSR---TGTPPGYRGTSPPAGCTFKDGS 683
|||:|||:|||:|||:
QY 158 PG-GISPVPSPGYPPNPSPGVGCYYPP-----GGYPATTSSQYPS----QPVTTVGPSRD 208
|||:|||:|||:|||:
Db 684 PTVGPCLP---PAGSGPLFSLPPPAPASPGLSATQIKQEPAAEYETESPVPFARS 740
|||:|||:|||:|||:
QY 209 GTIISEDITRASLISAVSDKLRWMKEEMDR-----AOELNALKRTREDLKKGHKQLDEM 263
|||:|||:|||:|||:
Db 741 FPPPKVVDPVSHASQA---RFNKHLDGFNSCARSDLYFPVEGSKLAKRADLVEK 796
|||:|||:|||:|||:
QY 264 VTRLDOEVAEDVNIELKKKDELSALEK 294
|||:|||:|||:|||:
Db 797 VEREAQRAREKEREREREKERERER 827
|||:|||:|||:|||:

RESULT 23
H81817
protein C18H9.8 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: H81817
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: H81817
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-678 <STO>
A:Cross-references: GB:chr_II; PID:g722387; GSPDB:GN00020; CESP:C18H9.8
A:Gene: C18H9.8
A:Map position: 2

Query Match 6.3%; Score 126.5; DB 2; Length 678;
Best Local Similarity 22.6%; Pred. No. 5.3;
Matches 84; Conservative 42; Mismatches 98; Indels 147; Gaps 18;

QY 93 VDANGKIYLPYLHEWKHP-----QSDILGLIQVMWIFGDEPPVFESRPISASY 140
|||:|||:|||:|||:
Db 32 VDDNNQMGLDYLTARDQGYKVITVAGVGHDLSGIA-----DKPEWSFOQNFF 82
|||:|||:|||:|||:
QY 141 ---PPYQAT-----GPPNTSYMFGMPGGISPYP---SGYPPNP-----SG 174
|||:|||:|||:|||:
Db 83 QLSPPQPPTYENRPTTCMSMENGPP---VPPSRSGMIPVPPSRSGGPPAPMPVSRRAG 138
|||:|||:|||:|||:
QY 175 YPCGPPPPGYPATTSSQYPSQ---PPVTTVG-----PSRDGT-----TSEDITRASLISA 223
|||:|||:|||:|||:
Db 139 PPRAPTSMMG-RPMTGWARPPTAGLRPVTOQGLRAPPSRMGTGNSRQVDFKSYIVGLRA 197
|||:|||:|||:|||:
QY 224 VSDKLIR---WRMKEMDRAQALNALKRTEE-----DLKKGHOKI--- 260
|||:|||:|||:|||:
Db 198 KQNAIVEIRSMKEKRDGKIDRNELHAYESRASAQAEISDLQGLLDLNKIMEKIHNL 257
|||:|||:|||:|||:
QY 261 -----FEMTRLDQFEAVDVKNIEL----- 280
|||:|||:|||:|||:
Db 258 GDMSDIIELEASKMKEADEMSLAEEVENERRAKEEELQGLEIEVEEQKLENAVTHAMD 317
|||:|||:|||:|||:

	Query Match	6.2%	Score 125;	DB 2:	Length 2715;
	Best Local Similarity	30.3%;	Pred. No. 35;		
Matches	43; Conservative	8; Mismatches	43; Indels	48; Gaps	8;
QY	71 PYNPPICFVKPTSSMTIKTKGHVDANGKIYLPHLHEWKHPQSDLLGLIQWIVVFGDEPP	130			
Dd	1257 PHPPP-----PHSPHTAAQ---QAAGQ-----HQQHHPQHQPGL-----PGPP	1292			
QY	131 VFSRPIASVPYPQATGPNTSNMGMGGISPYPSGY-----PNPNSGY-----PGCPYP	181			
Dd	1293 -----PPQQOQGQQGQPPPSVGSGFPAPQQHGHGVQVPPSQHVRFPAAGAPYP	1342			
QY	182 PGG-----PYPATTSQYPSQP	198			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 38.4494 Seconds

(without alignments)
3118.303 Million cell updates/sec

Title: US-09-804-690-4

Perfect score: 2002

Sequence: 1 MVSFKYRLVRETNNVIT.....FOLRALMQARKTAGLSLDLY 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum Match 100%

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1996	99.7	390	4	Q9BUM5
2	1863.5	93.1	391	11	Q7TSE5
3	1862	93.0	392	13	Q9I8G8
4	1000.5	50.0	408	5	Q9VVA7
5	754.5	37.7	425	5	Q762S8
6	575	28.7	249	5	Q8MQZ0
7	522	26.1	114	13	Q7T2M2
8	455	22.7	398	10	Q9LHG8
9	427	21.3	379	4	Q8IX04
10	424.5	21.2	249	11	Q7TQD3
11	422	21.1	174	11	Q08761
12	391	19.5	368	10	Q9FFV6
13	371	18.5	177	4	Q96FF5
14	371	18.5	341	4	Q9NWX7
15	303	15.1	385	3	P87279
16	301.5	15.1	378	3	Q873M7

17	294	14.7	111	11	Q8BU96
18	200	10.0	296	3	Q8NIM6
19	169	8.4	376	3	P78998
20	153.5	7.7	148	4	Q9NTQ8
21	151.5	7.6	485	13	Q804G3
22	148.5	7.4	148	4	Q9NZ81
23	148.5	7.4	397	4	Q8NEZ2
24	148.5	7.4	397	4	Q86DL9
25	147	7.3	397	11	Q8CHS8
26	145.5	7.3	1130	4	Q8IZL8
27	144.5	7.2	746	4	Q9BU60
28	144.5	7.2	1052	4	Q96FT1
29	144	7.2	483	13	Q7T391
30	142	7.1	1021	4	O15451
31	141.5	7.1	503	11	Q921F1
32	141	7.0	247	10	Q945K9
33	141	7.0	463	11	Q922A2
34	139	6.9	415	11	Q8CCV9
35	139	6.9	1284	4	O15450
36	138.5	6.9	827	3	Q80130
37	138.5	6.9	838	3	Q871B7
38	137	6.9	526	13	Q804G4
39	137	6.8	463	11	Q8BP75
40	136	6.8	137	11	Q9CQJ5
41	136	6.8	1066	10	Q7XE90
42	135.5	6.8	437	10	Q9STN0
43	135.5	6.8	895	10	Q9LIW7
44	135.5	6.8	1105	11	Q8BMJ4
45	135	6.7	463	11	Q8VIN2
46	135	6.7	674	5	Q18106
47	134.5	6.7	809	13	Q7ZVN7
48	134	6.7	201	16	O86316
49	134	6.7	485	3	Q9UTP6
50	133	6.6	198	16	Q7U0B9
51	132.5	6.6	471	10	Q9LD31
52	132.5	6.6	671	16	Q9Z502
53	132	6.6	371	4	Q7Z429
54	131.5	6.6	900	10	Q94B77
55	131.5	6.6	900	10	Q9FIU0
56	131	6.5	101	10	Q8LCI8
57	131	6.5	472	3	O59907
58	131	6.5	539	13	Q918F3
59	131	6.5	1880	5	O18465
60	130.5	6.5	332	13	Q8QGD9
61	130.5	6.5	609	17	Q8TXA4
62	130.5	6.5	739	10	Q8L7F7
63	130	6.5	345	11	Q9ESF4
64	130	6.5	1647	4	Q9HBD3
65	130	6.5	1679	4	Q9HBD4
66	129.5	6.5	157	5	Q9WIK1
67	129.5	6.5	320	11	Q8CI25
68	129.5	6.5	358	10	Q942Z3
69	129.5	6.5	1116	3	Q9HGL2
70	129	6.4	506	3	Q96UP4
71	129	6.4	1388	4	Q86VH6
72	129	6.4	1419	4	Q86U88
73	129	6.4	1449	4	O94837
74	129	6.4	3326	12	Q7T591
75	128.5	6.4	262	13	Q90713
76	128.5	6.4	754	10	Q8S683
77	128.5	6.4	1006	11	Q62901
78	128.5	6.4	1032	10	Q7XXN3
79	128	6.4	139	10	Q9SIH8
80	128	6.4	604	5	O97339
81	128	6.4	612	4	O96JV2
82	128	6.4	1099	10	Q8W362
83	128	6.4	1099	10	Q7XDA1
84	128	6.4	3942	11	O88737
85	127.5	6.4	284	4	Q9UBV8
86	127.5	6.4	343	4	Q81VW7
87	127.5	6.4	1129	12	Q9QR71
88	127	6.3	966	3	Q01385
89	127	6.3	977	3	Q8X005

```
90 127 6.3 1191 4 Q86V38 Q86v38 homo sapien
91 126.5 6.3 754 13 Q8J177 Q8j177 xenopus lae
92 126.5 6.3 840 10 Q8S688 Q8s688 oryza sativ
93 126.5 6.3 949 10 Q41524 Q41524 triticum ae
94 126.5 6.3 976 12 Q9DUN0 Q9dun0 kaposi's sa
95 126.5 6.3 979 5 Q9W5A4 Q9wsa4 drosophila
96 126.5 6.3 1114 5 Q8MT16 Q8mt16 drosophila
97 126.5 6.3 1114 5 Q7YZA4 Q7yza4 drosophila
98 126.5 6.3 1153 5 Q8IRY0 Q8iry0 drosophila
99 126 6.3 192 5 Q9VKM5 Q9vkm5 drosophila
100 126 6.3 1047 4 Q9C0B4 Q9c0b4 homo sapien
```

ALIGNMENTS

```
RESULT 1
Q9BUM5 PRELIMINARY; PRT; 390 AA.
AC Q9BUM5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tumor susceptibility gene 101.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002487; AA02487.1; -.
DR GO; GO:0004840; Fubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; P;ubiquitin cycle; IEA.
DR InterPro; IPR008883; Tsg101.
DR InterPro; IPR000608; UBQ-conjugat.
DR Pfam; PF05743; Tsg101; 1.
DR SMART; SM00212; UBCC; 1.
SQ SEQUENCE 390 AA; 43910 MW; 96BB2A4FC22DF16A CRC64;

Query Match 99.7%; Score 1996; DB 4; Length 390;
Best Local Similarity 99.7%; Pred. No. 1.5e-123;
Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MVSQYKRYDLTVRETQVNTVITLYKDLKPVLDVSYVFDGSSRELMLNLTGTPVYRGNTYNI 60
Db 11 MVSQYKRYDLTVRETQVNTVITLYKDLKPVLDVSYVFDGSSRELMLNLTGTPVYRGNTYNI 70
Qy 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130
Qy 121 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGIGSPYSPGYPNPSPGPGCPY 180
Db 131 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGIGSPYSPGYPNPSPGPGCPY 190
Qy 181 PPGGYPATSSQYPSQPPVTVTGPSRDGTHISEDTIRASLISAVSDKLRWRKEMDRQA 240
Db 191 PPGGYPATSSQYPSQPPVTVTGPSRDGTHISEDTIRASLISAVSDKLRWRKEMDRQA 250
Qy 241 AELNALKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKKDELSALKMENQSE 300
Db 251 AELNALKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKKDELSALKMENQSE 310
Qy 301 NNDIDIVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 360
Db 311 NNDIDIVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRKQ 370
Qy 361 FQRLMOKARKTAGLSLDLY 380
Db 371 FQRLMOKARKTAGLSLDLY 390
```

```
RESULT 2
Q7TSE5 PRELIMINARY; PRT; 391 AA.
AC Q7TSE5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tumor susceptibility gene 101 protein.
GN TSG101.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F344;
RA Burgdorf S.; Scheidtman K.H.;
RT "Functional interaction between AATF, TSG101 and AR.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY293306; AAP45008.1; -.
SQ SEQUENCE 391 AA; 44096 MW; F6274A94DBD46CA3 CRC64;

Query Match 93.1%; Score 1863.5; DB 11; Length 391;
Best Local Similarity 92.4%; Pred. No. 7.8e-115;
Matches 352; Conservative 14; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MVSQYKRYDLTVRETQVNTVITLYKDLKPVLDVSYVFDGSSRELMLNLTGTPVYRGNTYNI 60
Db 11 MVSQYKRYDLTVRETQVNTVITLYKDLKPVLDVSYVFDGSSRELMLNLTGTPVYRGNTYNI 70
Qy 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130
Qy 121 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGIGSPYSPGYPNPSPGPGCP 179
Db 131 MIVVFGDEPPVFSRPIASYPYQATGPPNTSYMPGPGGIGSPYSPGYPNPSPGPGCP 190
Qy 180 YPPGYPATSSQYPSQPPVTVTGPSRDGTHISEDTIRASLISAVSDKLRWRKEMDR 239
Db 191 YPPGYPATSSQYPSQPPVTVTGPSRDGTHISEDTIRASLISAVSDKLRWRKEMDR 250
Qy 240 QAEINLAKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKKDELSALKMENQ 299
Db 251 QAEINLAKRTEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKKDELSALKMENQ 310
Qy 300 ENNDIDIVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 359
Db 311 ENNDIDIVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 370
Qy 360 QFQRLMOKARKTAGLSLDLY 380
Db 371 QFQRLMOKARKTAGLSLDLY 391

RESULT 3
Q918G8 PRELIMINARY; PRT; 392 AA.
AC Q918G8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tumor susceptibility protein 101.
GN TSG101.
OS Chelonia mydas caranigra (Green sea-turtle).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Chelonioidae; Cheloniidae; Chelonia.
OX NCBI_TaxID=8469;
RN [1]
RP SEQUENCE FROM N.A.
RC Yu Q., Lu Y., Nerurkar V.R., Yanagihara R.;
```


RT "Studies on the turtle tumor susceptibility gene, TSG; full-length
RT cDNA sequence, genomic structure analysis and role in green turtle
RT fibropapilloma."; <http://www.ncbi.nlm.nih.gov/nuclot/111000000.html>;
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF279276; AAF87776.1; -.
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA

```

DR      GO:0000012, P:conjugation cycle
DR      InterPro; IPR008883; Tsg101.
DR      InterPro; IPR00608; UBQ_conjugat.

```

DR Pfam; PF05743; Tsg101; 1.
DR SMART; SM00212; UBCC; 1.

SQ SEQUENCE 392 AA; 44314 MW; 25DE98A5116CF8EA CRC64;

Query Match 93.0%; Score 1862; DB 13; Length 392;

Best Local Similarity 92.7%; Pred. NO: 9.8e-115;
Matches 354; Conservative 15; Mismatches 11;

QV 1 MVS KYRDLTVRET VNVITLYKDLKPVLDSYVENDGSSREL MNLTG

[illegible]

Quesada, A. and J. A. Rodríguez. 2002. *El agua en Chile: una perspectiva crítica*. Santiago: Editorial Nueva Universidad.

QY 6I FICLWELDIIFINPFFICFVKPFISSMIKIGRHDANGRIILFITHHE

Db 71 PICLWLLDTYFPFNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHEH

Qy 121 MIVFGDEPPVFSRP-ISASYPYQATGPPNTSYMPGMPGGISYPSP

Db
131 MIVFGEPPVFSRPTISTSFQYQATGPPNTSYMPGMPGSGISPYFP

Qy 180 YPPGGYPATTSSQ-YPSQPPVTTVGPSRDGTISEDITRASLISAVS

Db
Dd
191 YPPGGFFPATTSQGHYTSQPPVTTVGPSRDGTISEDITIRASRISAVE
191 :
191 :

Qv 239 30AEI NAL KRTEFDI KKGHOKI FEEMVTRIDOEVAEVDKNIETLKKKKT

[illegible]

DD Z51 AQAELNALNKIEEDLNKNGHQNDSEMVIRKEDHEVAVEDRNIEEDLNKNN

QY 299 SENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVII

Db 311 SENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVII

QY
359 KQFQLRALMQARKTAGLSDLY 380

Db 371 KQFQLRALMQKARKTAGLSDL 392

RESULT 4
09WA7

ID	Q9VVA7	PRELIMINARY;	PRT;	408 AA.
AC	Q9VVA7	OCODE.		

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE CG9712 protein (Tumor suppressor protein 101).
GN TSG101 OR CG9712.

OS *Drosophila melanogaster* (Fruit fly).

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorphi
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Ptery-
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Ptery-

OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;

RN [1]
PP SEQUENCE FROM N. A.

RC STRAIN=Berkeley;
BY MDTY INT 00100000 P-EM-F 10701120

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne
RA MEDLINE=20196006; PubMed=10731132;

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Henc
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Henc
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henc
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henc

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen
RA Brandon R.C., Rogers Y.-H.C., Blazer P.G., Chamne M.
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen

Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Nelson C.R., Rogers R.C., Stuej A.S., Champagne M., Strandon K.C.

RA Adria J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Ba

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bols

Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Baulke H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Daulton C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I.J., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupel M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 RL [2]
 RN
 RP
 RA SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Chame M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnicki F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RA "Annotation of *Drosophila melanogaster* genome.";
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RN
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN
 RP SEQUENCE FROM N.A.
 RA Laurencou A., Burtis K.C., Hawley S.;
 RT "TSG101 fly homolog.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003526; AAF49406.2; -.
 DR EMBL; AF315343; AAG29564.1; -.
 DR FlyBase; FBgn0036666; TSG101.


```

QY 353 VLLSRKQFQRLALMOKARK 372
DB 333 VLLSREQFFHRATAEKVRE 352

RESULT 13
Q96FF5 PRELIMINARY; PRT; 177 AA.
AC Q96FF5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011011; AH01011.1;
DR GO; GO:0004840; Fubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; Pubiquitin cycle; IEA.
DR InterPro; IPR000608; UBQ_conjugat.
DR SMART; SM00212; UBCC; 1.
KW Hypothetical protein.
SQ SEQUENCE 177 AA; 19552 MW; E733E710F8D0068A CRC64;

Query Match 18.5%; Score 371; DB 4; Length 177;
Best Local Similarity 60.0%; Pred. No. 5.7e-17;
Matches 63; Conservative 21; Mismatches 21; Indels 0; Gaps 0;

QY 29 LDSYFVNDGSSRELMLNLTGTPVYRGNTYNIPICLWLDTPYNPPICFVKPTSSMTIK 88
DB 1 MDITYVFKDSSQKDLLNFTGTPVMYQNTYNIPIRFLWILDSHPFAPPICFLKPTANMIL 60

QY 89 TGHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFS 133
DB 61 VGRHVDAGRIYLPYLQNWSHPKSVIVGLIKEMIAKFOELPMYS 105

RESULT 14
Q9NUX7 PRELIMINARY; PRT; 341 AA.
AC Q9NUX7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ11068 (EC 1.1.1.27) (L-lactate dehydrogenase) (LDH).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RL "NEDO human cDNA sequencing project.";
CC -|- CATALYTIC ACTIVITY: (S)-LACTATE + NAD(+) = PYRUVATE + NADH.
CC -|- PATHWAY: ANAEROBIC GLYCOLYSIS; FINAL STEP.
CC -|- SIMILARITY: BELONGS TO THE LDH FAMILY.
DR EMBL; AK001930; BA91985.1;
DR HSSP; P00336; SLDH.
DR GO; GO:0004459; F:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.

```

```

DR GO; GO:0006096; P:glycolysis; IEA.
DR GO; GO:0006512; P:ubiquitin cycle; IEA.
DR InterPro; IPR001236; LDH.
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR000205; NAD BS.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00056; ldh; 1.
DR Pfam; PF02866; ldh.C; 1.
DR PRINTS; PR00086; LDHHRGNASE.
DR SMART; SM00212; UBCC; 1.
KW Hypothetical protein; Glycolysis; NAD; Oxidoreductase.
SQ SEQUENCE 341 AA; 37587 MW; 7D5CD535296165FA CRC64;

Query Match 18.5%; Score 371; DB 4; Length 341;
Best Local Similarity 60.0%; Pred. No. 1.3e-16;
Matches 63; Conservative 21; Mismatches 21; Indels 0; Gaps 0;

QY 29 LDSYFVNDGSSRELMLNLTGTPVYRGNTYNIPICLWLDTPYNPPICFVKPTSSMTIK 88
DB 1 MDITYVFKDSSQKDLLNFTGTPVMYQNTYNIPIRFLWILDSHPFAPPICFLKPTANMIL 60

QY 89 TGHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFS 133
DB 61 VGRHVDAGRIYLPYLQNWSHPKSVIVGLIKEMIAKFOELPMYS 105

RESULT 15
P87279 PRELIMINARY; PRT; 385 AA.
ID P87279;
AC P87279;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE STP22P.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
ON NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=XGI#1;
RA Li Y., Kane T., Tipper C., Spatrick P., Jenness D.D.;
RL "The complete sequence of STP22 gene.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004731; AAB62820.1;
DR GO; GO:0005768; C:endosome; IDA.
DR GO; GO:0005515; F:protein binding; IDA.
DR GO; GO:0006612; P:protein-membrane targeting; IMP.
DR GO; GO:0006623; P:protein-vacuolar targeting; IMP.
DR InterPro; IPR008883; Tsg101.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF05743; Tsg101; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS00127; UBIQUITIN CONJUGAT 2; 1.
SQ SEQUENCE 385 AA; 43330 MW; FEDE3BE79F063BCE CRC64;

Query Match 15.1%; Score 303; DB 3; Length 385;
Best Local Similarity 26.2%; Pred. No. 4.4e-12;
Matches 102; Conservative 77; Mismatches 163; Indels 48; Gaps 15;

QY 7 YRD--LTVRETNNVITLYKDLKPLVDSYFVNDGSSRELMLNLTGTPVYRGNT--YNTIPIC 63
DB 25 YNDGRTTFHDSLALLDNFHSRLRPRTRVFTSHDGTPLQLLSITGTGSDGSSPHSIPVI 84

QY 64 LMLDTPYNPIC-----FVKPTSSMTIKTGHVDANGKIYLPYLHEWKHPQSDLLGL 117
DB 85 MWVPSMYVVPVPPFTISINLENFDMNTISSLPFIQYIDNSGWIAPILHCHWDPAAMNLMV 144

QY 118 IQVMIVVFGDEPPVFSRPIASAPPYQATGPTTSYMPGPGGISPYSPGYPNPSPGYG 177
DB 145 VOELMSLL-HEPPDQAPSLEPKENTQLQCEQNTPLPPPKPS----PHLKPPLP----- 194

QY 178 CPYPPGGYPATSSQYPSQPPVTVTVGFSRD-----GTISEDIRASLISAYSDKLRW 230

```

```
Db 195 ---PPPPQASNALDMDND-TDTPNHHMLQNLQTVVNNELFRED-VDIVADKILT 249
QY 231 R---MKEMDRAQAEINALKRTEDLKGHQKLEEMVTRLDQAEVAEVDKNNIELKKKDEE 287
Db 250 RQTVMQESTARPH-ETIAIDKNH--LRAVEQAIEQTMHSLNAQIDVLTAN----RAKVQQ 302
QY 288 LSSALEKMEQENNDIDEVIPTAPLYKQILNLYAEENAIETIFYLGEALRGVVIDLD 347
Db 303 FST-----SHVDEEDVNSIAVAKTDGLNQLNVAQDYALTDTIELSRMLHRGTIPLD 357
QY 348 VFLKHVRLSLRKQFQRLALMQARKTAGLS 377
Db 358 TVKQRELLARQQFLVRWHIQ--RITSPLS 395

RESULT 16
Q873M7 PRELIMINARY; PRT; 378 AA.
AC Q873M7;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE Vacuolar protein sorting 23.
GN VPS23.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA Blanchin-Roland S.; Gaillardin C.;
RT "New genes involved in genetic control of extracellular protease
RT synthesis in Yarrowia lipolytica.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Blanchin-Roland S.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AU509167; CAD48920.1; -.
SQ SEQUENCE 378 AA; 41788 MW; C5498CACCS5ED3669 CRC64;
```

```
Query Match 15.1%; Score 301.5; DB 3; Length 378;
Best Local Similarity 28.0%; Pred. No. 5.4e-12;
Matches 106; Conservative 46; Mismatches 139; Indels 87; Gaps 11;

QY 85 MTIKTGKHDANGKTYLPYLHEW-KHPQ-SDLLGLIQVMIVVFGDEPPVFSRPSISYPP 142
Db 1 MMLSFGNVDDNGRCYHEPYSBWSGSDPQNTNLATFLVLSDFSKPEPPYSRPTPEYGP 60
QY 143 -----YQATGPNTSYMPGMP----- 158
Db 61 PTYPSQNDQQQTHQMRPPLAPQSGMPTQPVTGQHAVPPVNRGTGHIQQQPMRTGFV 120
QY 159 --GGISPYPS-GYPNPSGYP-----GCPYPPGGYPATTSQYF---SQPPVTVGP- 205
Db 121 DLGGSSTHASEAVPLPKPPQPOWQGVPPQAQYQAQAHPLOASHHPQARGGPEL 180
QY 206 -----SRDGTISEDITRASLISA-----VSDKLRWRMKEEMDRAQAE 242
Db 181 SQQFTPHRSRASRTDIDMDTAKSSDEPAPKPPNPERMKALDHLQSKKEADAIGAN 240
QY 243 LNALKRTEDLKGHQKLEEMVTR-----LDQEAVIDKNNIELKKKDELSALEKMEQN 298
Db 241 VOADDAQIEGLWMKLSLEAGVSFVLLRLQLEAQADKNNKILSDKMDQARRVITQ-ARA 299
QY 299 SENNDIDEVIPTAPLYKQILNLYAEENAIETIFYLGEALRGVVIDLDVLKHVRLSLR 358
Db 300 CEIPDIDSAVCAENVFNQLVDLTAGEQAIDDTIYALSIALDREKITVEPFMKHVRNLAR 359
QY 359 KQFQRLALMQARKTAGL 376
Db 360 EKFIKVIATIDKIVAGAGL 377
```

```
RESULT 17
Q8BU96 PRELIMINARY; PRT; 111 AA.
AC Q8BU96;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Inferred; signaling molecule ATP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK086750; BAC39736.1; -.
DR GO; GO:0004840; Fubiquitin conjugating enzyme activity; IEA.
DR GO; GO:0006512; P-ubiquitin cycle; IEA.
DR InterPro; IPR000608; UBQ conjugat.
SQ SEQUENCE 111 AA; 12909 MW; 78EACCC8158EB92B3 CRC64;
```

```
Query Match 14.7%; Score 294; DB 11; Length 111;
Best Local Similarity 58.8%; Pred. No. 3.8e-12;
Matches 50; Conservative 16; Mismatches 19; Indels 0; Gaps 0;

QY 53 YRGNTYNIPICLWLLDTYPNPICFVKPTSMITKTGKHDANGKTYLPYLHEWKHPQS 112
Db 2 YQGYTNIPIRFWILDSPFADPICLFKPTANWEISVGKHDAGKRIYILPQLQNSHPKS 61
QY 113 DLLGLIQVMIVVFGDEPPVFSRPS 137
Db 62 AIVGLIKEMIAKFQELPLYSIPSS 86
```

```
RESULT 18
Q8NIM6 PRELIMINARY; PRT; 296 AA.
AC Q8NIM6;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein.
GN STP22.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92245758; PubMed=1574926;
RA Bolle P.A.; Gilliquet V.; Berben G.; Dumont J.; Hilger F.;
RT "Yeast sequencing reports: The complete sequence of K3B, a 7.9 kb
RT fragment between PGK1 on chromosome III, reveals the presence of seven
RT open reading frames.";
RL Yeast 8:205-213 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92244356; PubMed=1574125;
RA Oliver S.G.; der Aart O.J.; Agostoni-Carbone M.L.; Aigle M.;
RA Alberghina L.; Alexandraki D.; Antoine G.; Anwar R.; Ballesta J.P.;
RA Benit P.;
RT "The complete DNA sequence of yeast chromosome III [see comments].";
RL Nature 357:38-46 (1992).
```

RN [3] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=92327849; PubMed=1626432;
 RX Skala J., Purnelle B., Goffeau A.;
 RA "The complete sequence of a 10.8 kb segment distal of SUP2 on the
 RT right arm of chromosome III from *Saccharomyces cerevisiae* reveals
 RT seven open reading frames including the RVS161, ADP1 and POK genes."
 RL Yeast 8:409-417(1992).
 RN [4] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=92221691; PubMed=1561837;
 RX Benit P., Chanet R., Fabre F., Faye G., Fukuhara H., Sor F.;
 RA "Sequence of the sup61-RAD18 region on chromosome III of *Saccharomyces*
 RT *cerevisiae*."
 RL Yeast 8:147-153(1992).
 RN [5] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=92254506; PubMed=1580103;
 RX Wilson C., Bergantino E., Lanfranchi G., Valle G., Carignani G.,
 RA Frontali L.;
 RT "A putative serine/threonine protein kinase gene on chromosome III of
 RT *Saccharomyces cerevisiae*."
 RL Yeast 8:71-77(1992).
 RN [6] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=92254505; PubMed=1580102;
 RX Bateau N., Fremaux C., Hebrard S., Menara A., Aigle M., Crouzet M.;
 RA "The complete sequence of a 10.8kb fragment to the right of the
 RT chromosome III centromere of *Saccharomyces cerevisiae*."
 RL Yeast 8:61-70(1992).
 RN [7] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=92245759; PubMed=1574927;
 RX Sor F., Cheret G., Fabre F., Faye G., Fukuhara H.;
 RA "Yeast sequencing reports: Sequence of the HMR Region on Chromosome
 RT III of *Saccharomyces cerevisiae*."
 RL Yeast 8:215-222(1992).
 RN [8] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=92397594; PubMed=1523889;
 RX Wilson C., Grisanti P., Frontali L.;
 RA "The complete sequence of a 6146 bp fragment of *Saccharomyces*
 RT *cerevisiae* chromosome III contains two new open reading frames."
 RL Yeast 8:569-575(1992).
 RN [9] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=92397595; PubMed=1523890;
 RX Scherens B., Messenguy F., Gigot D., Dubois D.;
 RA "The complete sequence of a 9,543 bp segment of the left arm of
 RT chromosome III reveals five open reading frames including glucokinase
 RT and the protein disulfide isomerase."
 RL Yeast 8:577-586(1992).
 RN [10] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC Jimenez A.;
 RA Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 RN [11] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC Louis E.J.;
 RA Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 RN [12] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=93377412; PubMed=8368009;
 RX

RA Valle G.;
 RT "TA-repeat Microsatellites are closely associated with ARS consensus
 RT sequences in Yeast chromosome III."
 RL Yeast 9:753-759(1993).
 RN [13] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=94031061; PubMed=8266725;
 RX Slonimski P.P., Brouillet S.;
 RA "A data-base of chromosome III of *Saccharomyces cerevisiae*."
 RL Yeast 9:941-1029(1993).
 RN [14] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=91377317; PubMed=1897318;
 RX Rad M.R., Lutzenkirchen K., Xu G., Kleinhans U., Hollenberg C.P.;
 RA "The complete sequence of a 11,953 bp fragment from CIG on chromosome
 RT III encompasses four new open reading frames."
 RL Yeast 7:533-538(1991).
 RN [15] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC MEDLINE=95373282; PubMed=7645349;
 RX Rodriguez-Cousino N., Lill R., Neupert W., Court D.A.;
 RA "Identification and initial characterization of the cytosolic protein
 RT Ycr7p."
 RL Yeast 11:581-585(1995).
 RN [17] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC Gromadka R.;
 RA Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [18] SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RA MIPS;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X59720; CAC42964.1; -.
 DR SGD; S0000514; STP22.
 DR GO; GO:0004840; F-ubiquitin conjugating enzyme activity; IEA.
 DR GO; GO:0006512; P-ubiquitin cycle; IEA.
 DR InterPro; IPR008883; Tsg101.
 DR InterPro; IPR000608; UBQ_conjugat.
 DR Pfam; PF05743; Tsg101; 1.
 DR SMART; SM00212; UBCG; 1.
 DR PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
 DR Hypothetical protein.
 KW
 SQ SEQUENCE 296 AA; 33304 MW; FB72992580BB41D3 CRC64;
 Query Match 10.0%; Score 200; DB 3; Length 296;
 Best Local Similarity 24.2%; Pred. No. 1.9e-05;
 Matches 72; Conservative 57; Mismatches 124; Indels 44; Gaps 13;
 QY 7 YRD--LTRETIVNVTIKDLKPLVDSYVNDGSSRELMLNLTGTPVYRGNT--YNTPIC 63
 DB 25 YNDGRITTFHDSLLALDNFHSRLRPETRVTFHSDGTPQLLLSYGTISTGDSGSPHSIPVI 84
 QY 64 LWLLDTYPNPPIC-----FVKPTSSMTIKTKHVDANGKIYLPYLHEWKHPQSDLLGL 117
 DB 85 MWFSMYFVKVPPFSINLENFDMNTISSSLPIQEVDSNGWIALPILHCWDPAAANLIMV 144
 QY 118 IQVMIVVFGDEPPVFSPIASYPYQATGPTTSPYMPGPGGSLSPSPGYPSPGYPG 177
 DB 145 VQELMSLL-HEPPQDQAPSLPPKPNLTQLQQQNTPLPLPKPKS----PHLKPPPLP- 194
 QY 178 CPYPGGPYPATTSQYPSQPPVTVTVGFSRD-----GTISEDTIRASLISAYSVDKLRW 230
 DB 195 ---PPPPPPQASNALDMLMDMN-TDISFTNHEHMLQNLQTVVNELYRED-VDYVADKILT 249

Qy	231	R---MKEBMDRAQAEINLAKTTEEDLKKGHQKLEEMVTRLDOQVAEVDKNIELLKKK	284
Db	250	RQTYMQESIAERFH-EIIA-----IDKNHLRAVEQA--IEQTHSLNAQIDVLNRK	296
RESULT 19			
PR7898		PRELIMINARY;	
ID	P78998	PRT; 376 AA.	
AC	P78998;		
DT	01-MAY-1997 (TrEMBLrel. 03, Created)		
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein.		
GN	LGICLUN5C.		
OS	Saccharomyces pastorianus (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=27292;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Andersen T., Nilsson-Tillgren T.;		
RL	Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.		
EMBL	Z86109; CAB06793.1; ..		
GO	GO:0005622; C:intracellular; IEA.		
DR	GO:0005840; C:ribosome; IEA.		
DR	GO:0003735; F:structural constituent of ribosome; IEA.		
DR	GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.		
DR	GO:0006412; P:protein biosynthesis; IEA.		
DR	GO:0006512; P:ubiquitin cycle; IEA.		
DR	InterPro; IPR008195; Ribosomal_L34E.		
DR	InterPro; IPR008883; Tsg101.		
DR	InterPro; IPR000608; Ubq conjugat.		
DR	SMART; Pf05743; Tsg101; 1.		
DR	Swat; SM00212; UBCC; 1.		
DR	PROSITE; PS01145; RIBOSOMAL_L34E; 1.		
DR	PROSITE; PS01027; UBIQUITIN_CONJUGAT_2; 1.		
KW	Hypothetical protein.		
SQ	SEQUENCE 376 AA; 41542 MW; 837F91DA97BF0A0B CRC64;		
Query Match			
Best Local Similarity 8.4%; Score 169; DB 3; Length 376;			
Matches 60; Conservative 65; Mismatches 140; Indels 50; Gaps 9;			
Qy	11	TVRETVAVITLYKDLKPVLDSYFVNDGSSRELMMNLGTGPVYRGNTVNIPICLWLDDTY	70
Db	31	TFHDSALLDKFHLRLPRTRVTFTHSDGSPQLLSYGTGV-----GD--SLPLMIWIPSLY	83
Qy	71	PYNPPI-----CVKPTSSMTIKTGKHDVANGKIVLPYLHEWKHPQSDLLGLIQVMIV	124
Db	84	PVKRPFISIDLETDFVSAISSLSLPQAYIDSDGRVALPILDHWNPATNGLLIVVQELMSL	143
Qy	125	FGDEPPVFSRPIASYPYQATGPNTSYMPGMPGGISPYPSGYPPNPSPGYGCPYPPGG	184
Db	144	LSE-----PSRDVPSLP-----PKPDM-HSPLKHVTVPPLPPKA	177
Qy	185	PYPATTSOYPSQPPVTVGPS-----RDGTISEDTIRASLSAVSDKLRWRKKEMD	237
Db	178	KFPHV---QPPLQPPPPQPPSSAVDMLMDMDNTDLSPTNHHMLQLQSVLNELYREDVH	234
Qy	238	R-AQAELNALKRTEDLKKGHQKLEEMVTRLDOQVAEVDKNIELLKKKDELSALEKME	296
Db	235	YVADKILTRQIMQDSVARFHEMVAVDKTRLQAVEQTIEQMHTLNAQIEVLTAEATVQ	294
Qy	297	NQSENNDIDEVILPT 311	
Db	295	EFSTSPFDEEDVDT 309	
RESULT 20			
Q9NTQ8		PRELIMINARY;	
ID	Q9NTQ8	PRT; 148 AA.	
AC	Q9NTQ8;		

```
Query Match          7.6%; Score 151.5; DB 13; Length 485;
Best Local Similarity 35.5%; Pred. No. 0.055;
Matches 33; Conservative 6; Mismatches 33; Indels 21; Gaps 5;

QY 126 GDEPPV---PSRISASYPYQATGPNNTSYMPGPGISFP-----SGYP 169
DB 11 GSYPPASGPYQOP-AGYPPQAGYPPQAGYPPQAGYPPQAGYPPQAGYPPGAGYP 69

QY 170 PNPSPGPGCP---YPP-GGYPATTSQYPSQP 198
DB 70 PQAGGYPAAPGGGPPQAGGYPAAPGAYPNMP 102

RESULT 22
Q9NZ81 PRELIMINARY; PRT; 148 AA.
AC Q9NZ81;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Uncharacterized bone marrow protein BM041 (DKFZp564J157 protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Bone marrow.
RC TISSUE=B-cell;
RA Zhao M., Gu J., Li N., Peng Y., Han Z., Chen Z.;
RT "A novel gene expressed in human bone marrow.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217517; AAF67628.1; -
DR EMBL; BC016064; AAHL6064.1; -
DR EMBL; BC014257; AAHL4257.1; -
SQ SEQUENCE 148 AA; 15385 MW; 56BFE6A15935A2E CRC64;

Query Match          7.4%; Score 148.5; DB 4; Length 148;
Best Local Similarity 29.7%; Pred. No. 0.021;
Matches 41; Conservative 18; Mismatches 46; Indels 33; Gaps 6;

QY 126 GDEPPVPSRISASYPYQATGPNNTSYMPGPGISFPYPPNP---SGYPCG--- 178
DB 22 GGSNFAHPPTINPPPPGCPPPP-----PGAPHGNAFPFGGPHFPVQPGYGCQPLG 75

QY 179 PYPPGGYPATTSQYPSQPPVTVGSPRDTISEDITIRASLISAVSDKLRWRMKEMDR 238
DB 76 PYPPYPYPPA-----PGIPPNPLAPGMVGP-----AVIVDK---XMQKKKK 115

QY 239 QAELNALKRTPEDLKKG 256
DB 116 AHKKWKHKQHKHYKHG 133

RESULT 23
Q8NEZ2 PRELIMINARY; PRT; 397 AA.
AC Q8NEZ2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
```

```
RA Xu Z., Liang L., Zhao M., Li T.;
RT "A novel human gene in 8p22.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY033079; AAK54349.1; -
KW Hypothetical protein.
SQ SEQUENCE 397 AA; 44314 MW; 96EBB670F04A0923 CRC64;

Query Match          7.4%; Score 148.5; DB 4; Length 397;
Best Local Similarity 22.7%; Pred. No. 0.068;
Matches 84; Conservative 59; Mismatches 130; Indels 97; Gaps 17;

QY 51 VPYRGNTYNIPICLWLIDTYPNPPICFVKP--TSSMTIKTKGHVDANGKIYLPYLHEWK 108
DB 54 LPFTINNTINILPPQPPQKPVLSVYPPIRHLLMDKQGVVITS-----PLVNFT 107

QY 109 HPOSDDLGLIQVMIVFGDEPPVPSRISASYPYQATGPNNTSYMPGPGISFPYS-G 167
DB 108 -MHSDLGKIISLLDFWKNPPVLA-PTSTAFP-----YLYSNFSGMSPYASQG 154

QY 168 YPNPSPGPGCPYPP-----GGYPATTSQYPSQPPVTVGSPRDTISEDIT 215
DB 155 FPFLP-----PYPPQENRSITSLSVADTVSSSTSHHTAKPAASFGVLSNLPDPT 208

QY 216 IRASLISAVSDKLRWRMKEMDR-----AQAEALNALKRTPEDLKKGHQKLEEMVT--R 266
DB 209 VDAS-IPTSQNGFGYKMPDVPDAFPELSLSVSLQTDNMNEQEVL-----LEQFLTL 261

QY 267 LDQEVAVD---KNIELLKKDELSALE-----KMNQSEN 301
DB 262 LKQIITDKDLVKSIEELARKNLLPELSLEAKRQTVLDKYELLTQMKSTFEKKMQRQHEL 321

QY 302 NDIDEVIITAPLYKQILNLYAENAJEDTIFYLGEALRGVIDLDVFLK-----H 352
DB 322 SE-----SCSASALQARLKVAHH--AEESDNIAEDFLEKWEIDDFLSFMFKRICH 374

QY 353 VRLLSRKQFQ 362
DB 375 CRRAKBKLQ 384

RESULT 24
Q96DL9 PRELIMINARY; PRT; 397 AA.
AC Q96DL9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ32642.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovium;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK057204; BAB71381.1; -
KW Hypothetical protein.
SQ SEQUENCE 397 AA; 44342 MW; 7FF33F70B2EF4EC1 CRC64;

Query Match          7.4%; Score 148.5; DB 4; Length 397;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 9.98686 Seconds
(without alignments)
1981.269 Million cell updates/sec

Title: US-09-804-690-4
Perfect score: 2002
Sequence: 1 MVSKYKRDLTRETVNVT.....FOLRALMQARKTAGLSLDLY 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2002	100.0	390	1	T101_HUMAN
2	1900.5	94.9	391	1	Q98166 homo sapien
3	303	15.1	385	1	ST22_YEAST
4	160	8.0	466	1	ANX7_HUMAN
5	142.5	7.1	263	1	Y737_DROME
6	142	7.1	505	1	ANXB_HUMAN
7	141.5	7.1	503	1	ANXB_MOUSE
8	141	7.0	463	1	ANX7_MOUSE
9	138.5	6.9	4903	1	MLL3_MOUSE
10	136.5	6.8	503	1	ANXB_FABIT
11	130.5	6.5	338	1	FOSB_HUMAN
12	130	6.5	1647	1	SN24_HUMAN
13	128	6.4	788	1	PCAP_HUMAN
14	128	6.4	1078	1	DAM1_HUMAN
15	127	6.3	426	1	SUM2_SCHPO
16	127	6.3	1185	1	DRPL_HUMAN
17	126	6.3	868	1	PG61_HUMAN
18	126	6.3	1453	1	CA11_MOUSE
19	125	6.2	401	1	PD61_RAT
20	125	6.2	869	1	PD61_MOUSE
21	125	6.2	2716	1	OSA_DROME
22	124	6.2	964	1	YOY1_CAELI
23	123.5	6.2	379	1	VASP_HUMAN
24	123.5	6.2	2779	1	LVA_DROME
25	123	6.1	1248	1	DIA1_HUMAN
26	122	6.1	978	1	RA50_AQUAE
27	120.5	6.0	245	1	CM32_HUMAN
28	120.5	6.0	375	1	DRPL_MOUSE
29	120.5	6.0	753	1	ZIN_HUMAN
30	120	6.0	760	1	ZIN_MOUSE
31	120	6.0	1101	1	DIA2_HUMAN
32	119.5	6.0	503	1	ANXB_BOVIN
33	119.5	6.0	1224	1	DYNA_CHICK

34	119	5.9	1183	1	DRPL_RAT	P54258 rattus norv
35	119	5.9	1280	1	DYNA_RAT	P28023 rattus norv
36	118.5	5.9	383	1	VASP_CANFA	P50551 canis famli
37	118.5	5.9	792	1	PCAP_MOUSE	Q02482 mus musculu
38	118.5	5.9	1255	1	DIAL_MOUSE	O08808 mus musculu
39	118.5	5.9	1281	1	DYNA_MOUSE	O08788 mus musculu
40	118	5.9	535	1	Y475_STRMU	Q8dvk7 streptococc
41	118	5.9	1171	1	DIA3_MOUSE	O92507 mus musculu
42	118	5.9	1460	1	CA11_CANFA	O9xsi7 canis famli
43	117.5	5.9	620	1	EXTN_TOBAC	P13983 nicotiana t
44	117.5	5.9	1507	1	SET2_CAELI	Q18221 caenorhabdi
45	117	5.8	505	1	WASL_BOVIN	Q95107 bos taurus
46	116.5	5.8	338	1	FOSB_MOUSE	P13346 mus musculu
47	116.5	5.8	859	1	MUTS_AQUAE	O66652 aquifex aeo
48	116.5	5.8	1790	1	USO1_YEAST	P25386 saccharomyc
49	116	5.8	462	1	ANX7_DICDI	P24639 dictyostell
50	116	5.8	504	1	OCIN_CHICK	Q10449 gallus gall
51	116	5.8	609	1	YSWI_YEAST	P38280 saccharomyc
52	116	5.8	1464	1	CALL_HUMAN	P02452 homo sapien
53	116	5.8	2167	1	SHK1_RAT	Q9wv48 rattus norv
54	115.5	5.8	261	1	PRP2_MOUSE	P05142 mus musculu
55	115.5	5.8	475	1	S3A2_MOUSE	O62203 mus musculu
56	115	5.7	240	1	PRA_MYCTU	O53426 mycobacteri
57	115	5.7	467	1	CBPA_DICDI	P35085 dictyostell
58	115	5.7	559	1	MUS1_HUMAN	Q92558 homo sapien
59	115	5.7	786	1	MUS2_CLOTE	Q891u1 clostridium
60	115	5.7	1068	1	DAM2_HUMAN	O86t65 homo sapien
61	114.5	5.7	262	1	WBP2_RAT	Q8r478 rattus norv
62	114.5	5.7	684	1	CA39_HUMAN	Q14050 homo sapien
63	114.5	5.7	817	1	VRP1_YEAST	P37370 saccharomyc
64	114.5	5.7	1017	1	KCH4_HUMAN	Q9uq05 homo sapien
65	114.5	5.7	1940	1	MYH3_HUMAN	P11055 homo sapien
66	114	5.7	505	1	WASL_HUMAN	O00401 homo sapien
67	114	5.7	864	1	WS14_MOUSE	Q99m23 mus musculu
68	114	5.7	1017	1	KCH4_RAT	Q9rlt9 rattus norv
69	113.5	5.7	347	1	JUNB_HUMAN	P17275 homo sapien
70	113.5	5.7	1418	1	CA12_HUMAN	P02458 homo sapien
71	113	5.6	242	1	TPM1_PODCA	P41114 podocoryne
72	113	5.6	668	1	SCEL_HUMAN	P51171 homo sapien
73	113	5.6	736	1	DVL2_MOUSE	Q60838 mus musculu
74	113	5.6	852	1	WS14_HUMAN	Q9np71 homo sapien
75	113	5.6	902	1	NFC4_HUMAN	Q14934 homo sapien
76	113	5.6	1638	1	BRM_DROME	P25439 drosophila
77	112.5	5.6	582	1	MNT_HUMAN	Q99583 homo sapien
78	112.5	5.6	645	1	BRH2_DROME	Q24256 drosophila
79	112.5	5.6	834	1	YNC4_CAELI	P34537 caenorhabdi
80	112.5	5.6	1278	1	DYNA_HUMAN	Q14203 homo sapien
81	112.5	5.6	1940	1	MYH3_RAT	P12847 rattus norv
82	112	5.6	333	1	CAP1_DICDI	P19198 dictyostell
83	112	5.6	467	1	REP1_CRIGR	P11414 cricetulus
84	112	5.6	1068	1	DAM2_MOUSE	Q80u19 mus musculu
85	112	5.6	1234	1	PIB3_MOUSE	P51432 mus musculu
86	112	5.6	1695	1	CA54_HUMAN	P29400 homo sapien
87	112	5.6	1970	1	REP1_HUMAN	P24928 homo sapien
88	112	5.6	1970	1	REP1_MOUSE	P08775 mus musculu
89	112	5.6	2441	1	CBP_MOUSE	P45481 mus musculu
90	111.5	5.6	242	1	CM32_MOUSE	Q8bh93 mus musculu
91	111.5	5.6	639	1	SP01_MOUSE	Q15637 homo sapien
92	111.5	5.6	653	1	SP01_MOUSE	Q64213 mus musculu
93	111.5	5.6	1049	1	CA13_BOVIN	P04258 bos taurus
94	111.5	5.6	1550	1	ACSL1_ACEXY	P21877 acetobacter
95	111.5	5.6	1790	1	SEPA_EMENT	P78621 emeritella
96	111.5	5.6	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
97	111	5.5	520	1	WASP_MOUSE	P70315 mus musculu
98	111	5.5	867	1	PD61_XENLA	Q9w6c5 xenopus lae
99	111	5.5	1084	1	MYSS_RABIT	P02562 oryctolagus
100	111	5.5	1386	1	ZAP3_MOUSE	Q9r017 mus musculu

ALIGNMENTS

RESULT 1

```
T101_HUMAN
ID T101_HUMAN STANDARD; PRT; 390 AA.
AC Q99816;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tumor susceptibility gene 101 protein.
GN TSG101.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97148696; PubMed=9019400;
Li L., Li X., Francke U., Cohen S.N.;
RT "The TSG101 tumor susceptibility gene is located in chromosome 11 band
RT p15 and is mutated in human breast cancer.";
Cell 88:143-154(1997).
[2]
RL ERRATUM.
RP MEDLINE=99080983; PubMed=9867424;
Li L., Francke U., Cohen S.N.;
RA Cell 93:660-660(1998).
[3]
RN ALTERNATIVE SPLICING (ISOFORM 2).
RX MEDLINE=98031856; PubMed=9366528;
Gayther S.A., Barski P., Batley S.J., Li L., de Foy K.A., Cohen S.N.,
RA Ponder B.A., Caldas C.;
RT "Aberrant splicing of the TSG101 and RHIT genes occurs frequently in
RT multiple malignancies and in normal tissues and mimics alterations
RT previously described in tumors.";
Oncogene 15:2119-2126(1997).
[4]
RN ALTERNATIVE SPLICING.
RX MEDLINE=97384943; PubMed=9242438;
Lee M.P., Feinberg A.P.;
RA "Aberrant splicing but not mutations of TSG101 in human breast
RT cancer.";
Cancer Res. 57:3131-3134(1997).
[5]
RN ALTERNATIVE SPLICING.
RX MEDLINE=99054675; PubMed=9840940;
Wagner K.-U., Dierissee P., Rucker E.B. III, Robinson G.W.,
RA Hennighausen L.;
RT "Genomic architecture and transcriptional activation of the mouse and
RT human tumor susceptibility gene TSG101: common types of shorter
RT transcripts are true alternative splice variants.";
Oncogene 17:2761-2770(1998).
[6]
RN INTERACTION WITH DMAP1.
RX MEDLINE=20347709; PubMed=10888872;
Rountree M.R., Bachman K.E., Baylin S.B.;
RA "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at
RT replication foci.";
Nat. Genet. 25:269-277(2000).
[7]
RN STRUCTURE BY NMR OF 1-145.
RX MEDLINE=22001147; PubMed=12006492;
Pomilio O., Alam S.L., Rich R.L., Myszka D.G., Davis D.R.,
RA Sundquist W.I.;
RT "Structure and functional interactions of the Tsg101 UEV domain.";
EMBO J. 21:2397-2406(2002).
[8]
RN STRUCTURE BY NMR OF 1-145.
RX MEDLINE=22289677; PubMed=12379843;
Pomilio O., Alam S.L., Davis D.R., Sundquist W.I.;
RA "Structure of the Tsg101 UEV domain in complex with the PTAP motif of
RT the HIV-1 p6 protein.";
Nat. Struct. Biol. 9:812-817(2002).
CC -!- FUNCTION: May be involved in cell growth and differentiation and
CC act as a negative growth regulator.
CC
```

```
CC -!- SUBUNIT: Interacts with ubiquitin, stathmin and DMAP1.
CC -!- SUBCELLULAR LOCATION: Mainly cytoplasmic. Depending on the stage
CC of the cell cycle, detected in the nucleus.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist. Several shorter
CC isoforms are detected in primary breast cancers and other
CC tumors;
CC Name=1;
CC IsoId=Q99816-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q99816-2; Sequence=VSP_004440;
CC Note=Detected in normal as well as cancer tissues;
CC -!- TISSUE SPECIFICITY: Heart, brain, placenta, lung, liver, skeletal,
CC kidney and pancreas.
CC -!- DOMAIN: The UEV domain is required for the interaction of the
CC complex with ubiquitin.
CC -!- DOMAIN: The coiled coil domain may interact with stathmin.
CC -!- SIMILARITY: Contains 1 UEV (ubiquitin E2 variant) domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U82130; AAC2083.1; -.
DR PDB; 1KPP; 24-MAY-02.
DR PDB; 1KPO; 25-MAY-02.
DR PDB; 1MAP; 06-NOV-02.
DR PDB; 1MAQ; 06-NOV-02.
DR Genew; HGNC:15971; TSG101.
DR MIM; 601387; -.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR GO; GO:0006512; P:ubiquitin cycle; TAS.
DR InterPro; IPR008883; Tsg101.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF05743; Tsg101; 1.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; FALSE NEG.
DR PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; FALSE NEG.
DR Transprot; Protein transport; Ub1 conjugation pathway; Coiled coil;
KW Growth regulation; Alternative splicing; 3D-structure.
FT DOMAIN 1 133 UEV
FT DOMAIN 235 316 COILED COIL (POTENTIAL).
FT VARSPDIC 15 119 Missing (in isoform 2).
FT FTID=VSP_004440.
SQ SEQUENCE 390 AA; 43944 MW; ADD6912FC22DF162 CRC64;
Query Match 100.0%; Score 2002; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.2e-102;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVSQKYRDLTVRETNVITLYKDLKPVLDYVFNDSGSSRELMLNLTCTIPVYRGNTYNI 60
Db 11 MVSQKYRDLTVRETNVITLYKDLKPVLDYVFNDSGSSRELMLNLTCTIPVYRGNTYNI 70
Qy 61 PICLWLLDTPYPPNPPICFVKPTSMITKTGKVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
Db 71 PICLWLLDTPYPPNPPICFVKPTSMITKTGKVDANGKIYLPYLHEWKHPQSDLLGLIQV 130
Qy 121 MIVVFGDEPPVFRSPISASYPYQATGPPNTSYMPGMPGGISPYPSGYPNPPSGYPCPY 180
Db 131 MIVVFGDEPPVFRSPISASYPYQATGPPNTSYMPGMPGGISPYPSGYPNPPSGYPCPY 190
Qy 181 PPGGYPYATTSQYPSQPPVTVTVGSPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
Db 191 PPGGYPYATTSQYPSQPPVTVTVGSPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250
Qy 241 AELNALRTEEDILKKGKLEEMVTRLDQFVAEDKNIELLKKKDEELSSALEKMNQSE 300
```



```

Db      371 QFOLRALMQKARTAGLSLDY 391
|||||
RESULT 3
ST22 YEAST
ID ST22 YEAST STANDARD; PRT: 385 AA.
AC R25604; P87010; P87279; Q86ZT3; Q8NIM6;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Suppressor protein STP22 of temperature-sensitive alpha-factor
DE receptor and arginine permease (Vacuolar protein sorting-associated
DE protein VPS23).
GN STP22 OR VPS23 OR YCL008C OR YCL8C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RP MEDLINE=99223587; PubMed=10207082;
RX Li Y., Kane T., Tipper C., Spatrick P., Jenness D.D.;
RT "Yeast mutants affecting possible quality control of plasma membrane
RT proteins."
RL Mol. Cell. Biol. 19:3588-3599(1999).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=S288C;
RC MEDLINE=92244356; PubMed=1574125;
RX Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA Benit P., Berben G., Bergantinos E., Bileau N., Bolle P.-A.,
RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA Carignani G., Chanet R., Contreras R., Crouzet M., Daigman-Fornier B.,
RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA Francinguies-Gallard M.-C., Franco L., Frontali L., Fukuhara H.,
RA Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
RA Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haaseemann M.,
RA Hatat D., Hegemon J.H., Herbert C.J., Hilger F., Hohmann S.,
RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA Jacq C., Jaquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA Kleinhans U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,
RA Lucchini G., Lutzenkirchen K., Maat C., Mannheim G., Manzano M.E.,
RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA Perea J., Philippson P., Pierard A., Planta R.J., Plevani P.,
RA Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA Spiegelsberg R., Stave L.I., Steensma H.Y., Steiner S., Thierry A.,
RA Thieoes G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,
RA van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warrington J.R.,
RA von Wettstein D., Wickstead B.L., Wilson C., Wurst H., Xu G.,
RA Zimmermann F.K., Sgouros J.G.;
RT "The complete DNA sequence of yeast chromosome III."
RL Nature 357:38-46(1992).
RN [3]
REVISONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
REVISONS TO N-TERMINUS.
RA Valles G., Volckaerts G.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
SEQUENCE OF 204-329 FROM N.A.
RP STRAIN=S288C;
RC MEDLINE=22728591; PubMed=12844361;

```

```

RA Brachat S., Dietrich F.S., Voegel S., Zhang Z., Stuart L., Lerch A.,
RA Gates K., Gaffney T., Philippsen P.;
RT "Reinvestigation of the Saccharomyces cerevisiae genome annotation by
RT comparison to the genome of a related fungus: Ashbya gossypii."
RL Genome Biol. 4:RESEARCH45.1-RESEARCH45.13(2003).
RN [6]
FUNCTION, AND SUBCELLULAR LOCATION.
RP MEDLINE=21129038; PubMed=11208108;
RX Babst M., Odorizzi G., Estepa E.J., Emr S.D.;
RA "Mammalian tumor susceptibility gene 101 (TSG101) and the yeast
RT homologue, Vps23p, both function in late endosomal trafficking."
RL Traffic 1:248-258(2000).
RN [7]
FUNCTION, SUBUNITS, SUBCELLULAR LOCATION, AND MUTAGENESIS OF MET-85.
RP MEDLINE=21402413; PubMed=11511343;
RX Katzmann D.J., Babst M., Emr S.D.;
RA "Ubiquitin-dependent sorting into the multivesicular body pathway
RT requires the function of a conserved endosomal protein sorting
RT complex, ESCRT-I."
RL Cell 106:145-155(2001).
CC -|- FUNCTION: The ESCRT-I complex recognizes ubiquitinated
CC multivesicular body (MVB) cargo. It is required for sorting into
CC MVB vesicles. Required for vacuolar targeting of temperature-
CC sensitive plasma membrane proteins STS2 and CAN1.
CC -|- SUBUNIT: Component of ESCRT-I, endosomal sorting complex required
CC for transport I, which consists of STP22, VPS28 and SRN2.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic and endosomal.
CC -|- DOMAIN: The UEV domain is required for the interaction of the
CC complex with ubiquitin.
CC -|- SIMILARITY: Contains 1 UEV (ubiquitin E2 variant) domain.
CC -|- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 294.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004731; AAB62820.1; -.
CC EMBL; X59720; CAC42964.1; ALT_FRAME.
CC EMBL; AY260880; AAP21748.1; -.
CC PIR; S74288; S74288.
CC GERMOnline; 139851; -.
CC SGD; S0000514; STP22.
CC GO; GO:0005768; C:endosome; IDA.
CC GO; GO:0005515; P:protein binding; IDA.
CC GO; GO:0006612; P:protein-membrane targeting; IMP.
CC GO; GO:0006623; P:protein-vacuolar targeting; IMP.
CC InterPro; IPR008883; Tsg101.
CC InterPro; IPR000608; UEV_conjugat.
CC Pfam; PF05743; Tsg101; 1.
CC SMART; SM00212; UBCC; 1.
CC PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; FALSE_NEG.
CC PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
CC Transport; Protein transport; Ubl conjugation pathway; Coiled coil.
KW DOMAIN 87 164 UEV.
FT DOMAIN 155 201 COILED COIL (POTENTIAL).
FT DOMAIN 272 300 M-T: NO INTERACTION OF THE ESCRT-I
FT MUTAGEN 85 85 COMPLEX WITH UBIQUITIN.
SQ SEQUENCE 385 AA; 43330 MW; FEDE3BE79F063BCE CRC64;
Query Match 15.1%; Score 303; DB 1; Length 385;
Best Local Similarity 26.2%; Pred. No. 7.9e-10;
Matches 102; Conservative

```

```

Qy 7 YRD--LTVRETNNVITLYKDLKPVLDYSVFNDGSSRRLMNLGTIPVYPRGNT-YNIPIC 63
Db 25 YNDGRTTFHDSLLALDNFHSIRPRTRVFTHSDGTPLLSTGYTISTGDSGSSPHSIPVI 84

```


DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein CG31737 in chromosome 2.
 GN CG31737/CG18115 OR BG:DS02740.19.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RX STRAIN=Berkely;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=99403001; PubMed=10471707;
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R.G., Davis T.,
 RA Doyle C., Galle R.F., George R.A., Harris N.L., Hartzell G.,
 RA Harvey D., Hong L., Houston K.A., Hoskins R.A., Johnson G., Martin C.,
 RA Moshrefi A., Palazzolo M., Reese M.G., Spradling A.C., Tsang G.,
 RA Wan K.H., Whitelaw K., Celniker S.E., Rubin G.M.;
 RA "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region.";
 RL Genetics 153:179-219(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchman M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclebb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE003415; AAF45009.1; ALT_SEQ.
 CC EMBL; AE003650; AAF53528.3; -.
 CC FlyBase; FBgn0051737; CG31737.
 CC Hypothetical protein.
 FT DOMAIN 146 255 GLY/PRO-RICH.
 SQ SEQUENCE 263 AA; 28135 MW; 571C2860CCFF78D0 CRC64;
 Query Match 7.1%; Score 142.5; DB 1; Length 263;
 Best Local Similarity 24.6%; Pred. No. 0.26;
 Matches 58; Conservative 17; Mismatches 74; Indels 87; Gaps 9;
 QY 9 DLTVRTVNVITLYKDLKPVLDSYVF-----NDGSSRELMLNLTGTIPVYRGNT 57
 DB 64 DVNERHTN---HYKXHQSNPDEIATPNRSGVSSPNDGASSTOR-GTTTAPQGGKG 119
 QY 58 YNIPICLWLDDTYPNPPICFVKPTSSMTIKGKVDANGKIYLPVLHKKHPQSDLLGL 117
 DB 120 GN-----SPSRVTAQPPQPTI----- 136
 QY 118 IQVMIVVFGDEPVSFRSPISASVPPVQATGPPNNTSM-PMPGCGISPYPSGYP----- 169
 DB 137 --LLIVNKNDDPPYGGQSWVEYFGQYVFGPRGYGPGPRGYGPGPRGYGPRGRGY 194
 QY 170 --PNPSGPG-----CPYPPGPGVPATISS-----OYSPPPVTTVGPS 206
 DB 195 PGPGPRGYGPGPRYSCPGPRGYGPGSGRPPDGGGLQGYYPFGPGNGGTGPS 250
 RESULT 6
 ANXB HUMAN STANDARD; PRT; 505 AA.
 ID ANXB_HUMAN STANDARD; PRT; 505 AA.
 AC P50995; (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50)
 DE (56 kDa autoantigen).
 DE ANXA11 OR ANXA11.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 ON NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RP TISSUE=Teratocarcinoma;
 RX MEDLINE=94140847; PubMed=7508441;
 RA Misaki Y., Pruijn G.J.M., van der Kemp A.W., van Venrooij W.J.;
 RT "The 56K autoantigen is identical to human annexin XI.";
 RL J. Biol. Chem. 269:4240-4246(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20469408; PubMed=11013079;
 RA Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
 RA Fernandez M.P.;
 RT "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
 RT annexins and source of orthologous cDNA isoforms.";
 RL Genomics 69:95-103(2000).
 RN [3]

```
SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND POSSIBLY NUCLEAR.
CC -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC calcium and phospholipid.
CC -!- DISEASE: Antibodies against ANXA1 are present in sera from
CC patients with various autoimmune diseases, predominantly in sera
CC from patients with rheumatoid arthritis, systemic lupus
CC erythematosus, or Sjogren's syndrome.
CC -!- SIMILARITY: Belongs to the annexin family.
CC -!- SIMILARITY: Contains 4 annexin repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L19605; AAA19734.1; -.
CC EMBL: AJ278463; CAB94995.1; -.
CC EMBL: AJ278464; CAB94996.1; -.
CC EMBL: AJ278465; CAB94997.1; -.
CC EMBL: BC007564; AAH07564.1; -.
CC PIR: A53152; A53152.
CC HGSP: P13214; IAW.
CC Genew: HGNC:535; ANXA11.
CC MIM: 602572; -.
CC GO: GO:0005737; C:cytoplasm; TAS.
CC GO: GO:0005635; C:nuclear membrane; NAS.
CC GO: GO:0005654; C:nucleoplasm; NAS.
CC GO: GO:0005543; F:phospholipid binding; TAS.
CC GO: GO:0005515; F:protein binding; IPI.
CC GO: GO:0006955; P:immune response; TAS.
CC InterPro: IPR001464; Annexin.
CC Pfam: PF00191; annexin; 4.
CC PRINTS: PR00196; ANNEXIN.
CC ProDom: PD000143; Annexin; 4.
CC SMART: SM00335; ANX; 4.
CC PROSITE: PS00223; ANNEXIN; 4.
CC Annexin; Calcium/phospholipid-binding; Repeat; Polymorphism.
KW ANNEXIN 1.
FT REPEAT 209 269 ANNEXIN 1.
FT REPEAT 281 341 ANNEXIN 2.
FT REPEAT 365 425 ANNEXIN 3.
FT REPEAT 440 500 ANNEXIN 4.
FT VARIANT 230 230 R -> C (in dbSNP:1049550).
FT VARIANT 457 457 /FTId=VAR_012006.
FT VARIANT 457 457 I -> V (in dbSNP:1802932).
FT /FTId=VAR_012007.
SQ SEQUENCE 505 AA; 54389 MW; 4ADCAC8F270BFEE4 CRC64;
```

Query Match 7.1%; Score 142; DB 1; Length 505;
Best Local Similarity 28.0%; Pred.No. 0.59;
Matches 63; Conservative 19; Mismatches 91; Indels 52; Gaps 10;

QY 111 QSDLLGLIQWIVVFG--DEPPVSRPISASPPYQ--ATGPPNTSYMPGPGGIGSPYPS 166
Db 51 QDYLSGWAANMSGTFFGANMNLYPGAPGAGYPPVPPGGFGQPPSAQQQFVPPYGMPPPG 110
QY 167 GYPN--PS--GYPCGYP-----PGPYPATSSQYSPQPPVTTVG-----PSR 207
Db 111 GNPSRMSYPYPGAPVGGPMPPGQPPGAYGQPPVTPYGPVPLFGQQQPVESY 170
QY 208 DGTISEDTI-----RASLSAV-----SKLRWRKMEEMDRRAELNALKRT 249
Db 171 PGYSGGTVTFAVPTTGFSGRTTDAFGFDPLRDAEVLRYKAMRGFGTDEQAIIDCLGR 230
QY 250 EEDLKKGHQKLEEMVTRLDQVAEVDKNIELLLKKKDEELSSALEK 294
Db 231 SN--KQKQILLSPFTAYGK-----DLIKDKSELSGNFEK 264

RESULT 7
ANXB MOUSE
ID -ANXB MOUSE STANDARD; PRT; 503 AA.
AC P97384;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
GN ANXA11 OR ANX11
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP -SEQUENCE FROM N.A.
RX MEDLINE=97092887; PubMed=8938449;
RX Fernandez M.P., Jenkins N.A., Gilbert D.J., Copeland N.G.,
RA Morgan R.O.;
RT "Sequence and chromosomal localization of mouse annexin XI";
RL Genomics 37:366-374(1996).
[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=20469408; PubMed=11013079;
RA Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
RA Fernandez M.P.;
RT "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
annexins and source of orthologous cDNA isoforms";
RL Genomics 69:95-103(2000).
CC -!- FUNCTION: Binds specifically to calcyclin in a calcium-dependent
manner.
CC -!- DOMAIN: A pair of annexin repeats may form one binding site for
calcium and phospholipid.
CC -!- SIMILARITY: Belongs to the annexin family.
CC -!- SIMILARITY: Contains 4 annexin repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U65986; ABA42012.1; -.
CC EMBL: AJ289760; CAB94770.1; -.
CC EMBL: AJ289761; CAB94770.1; JOINED.
CC EMBL: AJ289762; CAB94770.1; JOINED.
CC EMBL: AJ289763; CAB94770.1; JOINED.
CC EMBL: AJ289764; CAB94770.1; JOINED.
CC EMBL: AJ289765; CAB94770.1; JOINED.
CC EMBL: AJ289766; CAB94770.1; JOINED.

[illegible]

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L49169; AAB53946.1; -.
 PIR; I53043; I53043.
 HSSP; P01100; LFOS.
 Genew; HGNC:3797; FOSE.
 MIM; 164772; -.
 GO; GO:0003677; F:DNA binding; TAS.
 GO; GO:0008134; F:transcription factor binding; TAS.
 GO; GO:0007610; P:behavior; TAS.
 GO; GO:0007275; P:development; TAS.
 GO; GO:000122; P:negative regulation of transcription from P. .; TAS.
 InterPro; IPR000837; Leuzip_Fos.
 InterPro; IPR004827; TF_bZIP.
 Pfam; PF00170; bZIP; 1.
 PRINTS; PR00042; LEUZIPRPOS.
 SMART; SM00338; BRLZ; 1.
 PROSITE; PS0217; bZIP; 1.
 PROSITE; PS00036; bZIP_BASIC; 1.
 Nuclear protein; DNA-binding.
 DNA BIND 161 179 BASIC MOTIF.
 DOMAIN 183 211 LEUCINE-ZIPPER.
 CONFLICT 338 338 L -> R (IN REF. 2).
 SEQUENCE 338 AA; 35927 MW; DDFE827C5047850F CRC64;
 Query Match 6.5%; Score 130.5; DB 1; Length 338;
 Best Local Similarity 28.3%; Pred. No. 1.6;
 Matches 51; Conservative 18; Mismatches 68; Indels 43; Gaps 7;
 QY 113 DLLGLIQWIV-----VGDPEPPVPSRISASYPYQATGPNPTSY-MPGMPGGIS 162
 Db 66 DLQWLIVQTLTSSMAQSQGQPLASQPPVVD-----PYDM---PGTSYSTPGMSGYSS 114
 QY 163 PYPSPYPPNPGSGPCYPYPPGYPATTSQVPSQPPVTTVPGRSDGTISEDTIRASLIS 222
 Db 115 GGASG-----SGPSTSGTSGPGPARPARPREETITPEE---EKRRVRERNK 166
 QY 223 AVSKLWRMKEMDRQAELNALKRTEDLKGHKLEEMVRLDQBAEVDKNIELLK 282
 Db 167 LAAAKCRNRRELIDRLQAETD-----QLEEEKAELESEIAELQKEKERLE 212
 RESULT 12
 ID SN24 HUMAN STANDARD; PRT; 1647 AA.
 AC P51532;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Possible global transcription activator SNF2L4 (SNF2-beta) (BRG-1 protein) (Mitotic growth and transcription activator) (Brahma protein homolog 1).
 DE SMARCA4 OR SNF2L4 OR BRG1 OR SNF2B.
 GN Homo sapiens (Human)
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94050144; PubMed=8232556;
 RA Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
 RT "BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription.";
 RL Nature 366:170-174(1993).
 RP [2]
 RP REVIEWS.
 RA Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94268902; PubMed=8208605;

RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
 RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
 RT Drosophila brahma are transcriptional coactivators cooperating with
 RT the estrogen receptor and the retinoic acid receptor.";
 RL Nucleic Acids Res. 22:1815-1820(1994).
 RN [4]
 RP SEQUENCE OF 814-1474 FROM N.A.
 RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
 RA Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Stilwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 RA Danganan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Attix C., Andreise I., Frankheim M., Amico-Keller G., Coefield J.,
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
 RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Carrano A.V.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP INTERACTION WITH NR3C1 AND PGR.
 RX MEDLINE=98250578; PubMed=9590696;
 RA Fryer C.J., Archer T.K.;
 RT "Chromatin remodelling by the glucocorticoid receptor requires the
 RT BRG1 complex.";
 RL Nature 393:88-91(1998).
 RN [6]
 RP IDENTIFICATION IN THE BAF53 COMPLEX WITH BAF53A; RUVBL1 AND TRRAP.
 RX MEDLINE=21829261; PubMed=11839798;
 RA Park J., Wood M.A., Cole M.D.;
 RT "BAF53 forms distinct nuclear complexes and functions as a critical
 RT c-Myc-interacting nuclear cofactor for oncogenic transformation.";
 RL Mol. Cell. Biol. 22:1307-1316(2002).
 CC -!- FUNCTION: Transcriptional coactivator cooperating with nuclear
 CC hormone receptors to potentiate transcriptional activation.
 CC -!- SUBUNIT: Interacts with NR3C1 and PGR. Component of the BAF53
 CC complex, at least composed of BAF53A, RUVBL1, SMARCA4/BRG1, and
 CC TRRAP, which preferentially acetylates histone H4 (and H2A) within
 CC nucleosomes.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; U29175; AAB40977.1; -.
 EMBL; D26156; BAA05143.1; -.
 EMBL; AC006127; AAC97987.1; -.
 PIR; S45252; S45252.
 TRANSFAC; T05451; -.
 Genew; HGNC:11100; SMARCA4.
 MIM; 603254; -.
 GO; GO:0005634; C:nucleoplasm; TAS.
 GO; GO:0004386; F:helicase activity; TAS.
 GO; GO:0003713; F:transcription co-activator activity; TAS.
 GO; GO:0003700; F:transcription factor activity; TAS.
 GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
 InterPro; IPR006576; BRK.
 InterPro; IPR001487; Bromodomain.
 InterPro; IPR001410; DEAD.
 InterPro; IPR001650; Helicase_C.
 InterPro; IPR006562; HSA.
 InterPro; IPR000330; SNF2_N.
 Pfam; PF00439; Bromodomain_1.
 Pfam; PF00271; helicase_C; 1.
 Pfam; PF00176; SNF2 N; 1.
 PRINTS; PR00503; BROMODOMAIN.
 SMART; SM00592; BRK; 1.
 SMART; SM00297; BROMO; 1.
 SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELIC; 1.
 DR SMART; SM00573; HSA; 1.
 DR PROSITE; PS00633; BROMODOMAIN 1; 1.
 DR PROSITE; PS00114; BROMODOMAIN 2; 1.
 KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
 KW ATP-binding; Helicase.
 FT DOMAIN 578 588 POLY-LYS.
 FT DOMAIN 663 672 POLY-GLU.
 FT NP_BIND 779 786 ATP (POTENTIAL).
 FT SITE 881 884 DEGH BOX.
 FT DOMAIN 1360 1364 POLY-GLU.
 FT DOMAIN 1477 1547 BROMODOMAIN.
 FT DOMAIN 1571 1584 POLY-GLU.
 FT SEQUENCE 1647 AA; 184585 MW; 78785E7953277FID CRC64;
 Query Match 6.5%; Score 130; DB 1; Length 1647;
 Best Local Similarity 21.7%; Pred. No. 10;
 Matches 61; Conservative 27; Mismatches 71; Indels 122; Gaps 11;
 QY 126 GDEPPVFSRISASYPYQATGPNSTSYMPCMP-----GGISYPSPG----- 167
 DB 243 GPAPPNYSRPHGMGPNPMPGP--SGVPEGMFGPPGPPKPPWPGPMANAAAPTSTPQ 300
 QY 168 --YPPNPSGYPGCPYPPGGP-----YPATTSSQVPSQPP-----VTVGPSPR 207
 DB 301 KLIPPOPTGRPS-PAPPAPVPPAASPVPVPPQTPGQPAQAPMVPPLHQKQSRITPIQKPR 359
 QY 208 -----DGTISED-----TI----- 216
 DB 360 GLDPVELQEREYRLQARIAHRIQELNPLGSLAGDLRTKATTELKALRLNLFQRLRQE 419
 QY 217 -----RASLISA-VSDKLRWPKMEEMDBA-----QAEINAL 246
 DB 420 VVVCMRDRDTALNNAKAYKRSQSLREARITEKLEKQKQTEQERKRRKQKQYLNLSI 479
 QY 247 KRTEEDLKQKHLEEMVTRLDQEVAVDKNIELLKKKDEE 287
 DB 480 LQHAQDFEYHRSVTGKIQLTKAVATYHANTEREQKENE 520
 RESULT 13
 PCAP_HUMAN STANDARD; BRT: 788 AA.
 AC Q6GRN5; O15413; Q8NF16; Q96CT0; Q96TH7; Q9P1T3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Positive cofactor 2 glutamine/Q-rich-associated protein (PC2
 DE glutamine/Q-rich-associated protein) (TPA-inducible gene-1) (TIG-1)
 DE (Activator-recruited cofactor 105 kDa component) (ARC105) (CTG repeat
 DE protein 7a).
 GN PCQAP OR TIG1 OR ARC105 OR CTG7A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 OX [1]
 SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Megakaryocytes, and Placenta;
 RX MEDLINE=20480707; PubMed=11024300;
 RA Abraham S., Solomon W.B.;
 RT "A novel glutamine-rich putative transcriptional adaptor protein
 RT (TIG-1), preferentially expressed in placental and bone-marrow
 RT tissues.";
 RL Gene 255:389-400(2000).
 RN [2]
 SEQUENCE FROM N.A. (ISOFORM 2), AND POLYMORPHISM OF POLY-GLN REGION.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=21303066; PubMed=11414760;
 RA Berti L., Mittler G., Przemeck G.K.H., Stelzer G., Guenzler B.,
 RA Amati F., Conti E., Dallapiccola B., Hrabe de Angelis M., Novelli G.,
 RA Meisternest M.;
 RT "Isolation and characterization of a novel gene from the Digeorge

RT chromosomal region that encodes for a mediator subunit.";
 RL Genomics 74:320-332(2001).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Hepatoma, and Spleen;
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Eye, Kidney, and Uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Spapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield A.C., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP SEQUENCE OF 185-573 FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain cortex;
 RX MEDLINE=97369492; PubMed=9225980;
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
 RT "CDNAs with long CAG trinucleotide repeats from human brain.";
 RL Hum. Genet. 100:114-122(1997).
 RN [6]
 RP IDENTIFICATION IN ARC COMPLEX, AND SEQUENCE OF 39-48 AND 525-536.
 RX MEDLINE=99249346; PubMed=10235267;
 RA Naeae A.M., Beaurang P.A., Zhou S., Abraham S., Solomon W.B.,
 RA Tjian R.;
 RT "Composite co-activator ARC mediates chromatin-directed
 RT transcriptional activation.";
 RL Nature 398:828-832(1999).
 CC -!- FUNCTION: May function as a transcriptional coactivator in RNA
 CC polymerase II transcription. As a protein complex component it may
 CC regulate gene transcription, especially following induction of
 CC protein kinase C activity.
 CC -!- SUBUNIT: Subunit of the large multiprotein complexes PC2 and
 CC ARC/DRIP.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q96RN5-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q96RN5-2; Sequence=VSP_003922;
 CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined, including
 CC heart, brain, lung, spleen, thymus, pancreas, blood leukocyte and
 CC placenta. However, the level of expression varied, with highest
 CC expression in the placenta and peripheral blood and lowest in the
 CC pancreas and kidney.
 CC -!- INDUCTION: By 12-O-tetradecanoylphorbol-13-acetate (TPA).
 CC -!- POLYMORPHISM: The poly-Gln region from amino acids 235-262 of
 CC PCQAP is polymorphic. There are from 15 to 18 repeats in the
 CC Italian population.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts


```
CC      in positions 13, 600 and 749.
CC      -!- CAUTION: Ref.3 (BAB85034) sequence differs from that shown due to
CC      sequencing errors.
CC
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; AF056191; AAC12944.1; ALT_FRAME.
DR      EMBL; AF328769; AAK58423.1; -.
DR      EMBL; AK074268; BAB85034.1; ALT_SEQ.
DR      EMBL; AK090465; BAC03446.1; ALT_INIT.
DR      EMBL; BC007529; AAH07529.1; -.
DR      EMBL; BC013985; AAH13985.1; -.
DR      EMBL; BC017110; AAH17110.1; ALT_INIT.
DR      EMBL; U07045; AAB91443.1; -.
DR      Genew; HGNC:14248; PCQAP.
DR      MIM; 607372; -.
KW      Transcription regulation; Activator; Nuclear protein; Polymorphism;
KW      Triplet repeat expansion; Alternative splicing.
FT      DOMAIN 547 564 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT      DOMAIN 161 174 POLY-GLN.
FT      DOMAIN 178 193 POLY-GLN.
FT      DOMAIN 205 218 POLY-GLN.
FT      DOMAIN 226 239 POLY-GLN.
FT      DOMAIN 243 262 POLY-GLN.
FT      DOMAIN 266 315 PRO-RICH.
FT      DOMAIN 360 367 POLY-GLN.
FT      DOMAIN 449 456 POLY-PRO.
FT      DOMAIN 602 611 MISSING (in isoform 2).
FT      VARSPLIC 385 424 /FTID=VSP_003922.
FT      VARIANT 261 262 Missing.
FT      CONFLICT 12 12 /FTID=VAR_013136.
FT      CONFLICT 116 116 L -> R (IN REF. 1).
FT      CONFLICT 154 154 Q -> H (IN REF. 1).
FT      CONFLICT 161 161 Q -> R (IN REF. 3; BAB85034).
FT      CONFLICT 185 186 QQ -> EL (IN REF. 5).
FT      CONFLICT 232 287 MISSING (IN REF. 3; BAB85034).
FT      CONFLICT 265 265 Q -> E (IN REF. 1 AND 5).
FT      CONFLICT 572 573 IL -> GI (IN REF. 5).
FT      CONFLICT 685 685 L -> V (IN REF. 3).
SQ      SEQUENCE 788 AA; 86753 MW; BB6AC6C63ED2F97E CRC64;

Query Match          6.4%; Score 128; DB 1; Length 788;
Best Local Similarity 23.4%; Pred.No.5.6;
Matches 63; Conservative 30; Mismatches 74; Indels 102; Gaps 15;

Qy 90 GKVDANGKIYLPYLHWEKHKQSDLLGLIQMIVVFGDEP-----PVFSRP----- 135
Db 394 GMHIRA-----RFPPTAVSAIPSSIPPLGQPWAQVQSQSILPMLSPSPCQQ 441
Qy 136 --ISASPPYQATGPNTSYMPGMP-----GGISPGSGYPPNPSPGPGCPYPPGG 184
Db 442 VQTQSMPP-----PPQSPQPGQPGSSQPNVSSGPPAPSPSSFLPSPS-----PQPSQS 491
Qy 185 PYPATISSQY--PSQPPVTT-VGSRDGTISEDITIRASLSIAVSKDLRWKKEMDRAQA 241
Db 492 PVTARTPQNFSPSPGPGPLNTPNPFS-----SWMSPAGSS-----QAE-----QQ 531
Qy 242 ELNALKRTEEDLKKGHKLEEMVTRLDQEAVDKNIIELLKKD----- 285
Db 532 YLDKLIKQ----LSXVIEPLREWINKIDK---NEDEKKDLKSKSLDILTDPSRCPLKT 584
Qy 286 -EELSSALEKMEKENSENDIDEVIPTAP 313
Db 585 LQKCEIALEKLN-----DMAVPTPP 605
```

```
RESULT 14
DAML HUMAN
ID DAML HUMAN STANDARD: PRT; 1078 AA.
AC Q9Y4D1; Q86U34; Q8N1Z8; Q8TB39;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Disveiled associated activator of morphogenesis 1.
GN DAAM1 OR KIAA0666.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=12508121;
RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
RA Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,
RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cruaud C.,
RA Bruels T., Jallou O., Friedlander L., Samson G., Brottier P.,
RA Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,
RA Aiach N., Boscus D., Dickhoff R., Dors M., Dubois I., Friedman C.,
RA Gouyvenoux M., James R., Madan A., Mairey-Estrada B., Mangenot S.,
RA Martins N., Menard M., Oztas S., Ratcliffe A., Shafer T., Trask B.,
RA Vacherie B., Bellemere C., Belser C., Resnard-Gonnet M.,
RA Bartol-Mavel D., Boutard M., Briez-Silla S., Combette S.,
RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Muselet D.,
RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukiewicz P., Trybou A.,
RA Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,
RA Dumont C., Guerin T., Haffray S., Hammadi R., Muanga J., Pellouin V.,
RA Robert J., Wunderle E., Gauguier G., Roy A., Sainte-Marthe L.,
RA Verdier F., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,
RA Matsuda F., Wilson R., Scarpelli C., Gypay G., Wincker P., Saurin W.,
RA Quetier F., Waterston R., Hood L., Weissenbach J.;
RT "The DNA sequence and analysis of human chromosome 14.";
RL Nature 421:601-607(2003).
RN [3]
RP SEQUENCE OF 1-497 FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal brain;
RA Li W.B., Gruber C., Jessee J., Polayes D.;
RT "Full-length cDNA libraries and normalization.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 706-1078 FROM N.A. (ISOFORM 3).
RC TISSUE=Thymus;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 407-1078 FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

FT	VARSPLIC	788	818		/FTid=VSP_008000.
FT	Missing	(in isoform 3).			
FT	VARSPLIC	888	916		/FTid=VSP_008001.
FT	NMLELKEIETSLRSGLKAVETELEVQKSK -> KSWNIIRSL				
FT	SPPSRISICLSASSGS (in isoform 3).				
FT	VARSPLIC	917	1078		/FTid=VSP_008002.
FT	Missing	(in isoform 3).			
FT	VARSPLIC	940	940		/FTid=VSP_008003.
FT	CONFLICT	1078 AA;	123472 MW;	E82D989239025486 CRC64;	
FT	SEQUENCE	1078 AA;	123472 MW;	E82D989239025486 CRC64;	
Query Match		6.4%;	Score 128;	DB 1;	Length 1078;
Best Local Similarity		23.3%;	Pred. No. 8;		
Matches	61;	Conservative	44;	Mismatches	117;
				Indels	40;
				Gaps	10;
QY	134	RPIGASYPPTATCGPNTSYMPMGPGIGISPYPSGYPNPSGV---	PGCFYPVGGVPATT	190	
Db	521	RAVCASIPGGSPGAPGGPFSSVPGSLLP-PPPPPPLPGMLPPPPPPLPGGPPPP-	577		
QY	191	SSQYPSQPVTTCVPSRDGTISEDITRASLISAVS-----	DCLRWMKEEMDR	238	
Db	578	---PGPPGLGAIMPGPAGPWGLAKKKSIPTNALKGFNWSKLPENKLEGVWTEIDD	633		
QY	239	AOA-----ELNALKTETEDLKGHQKLEEMVTRLDQEAIVVDKNIEL--	KXDEELSSL	292	
Db	634	TKVFKILDELERT----FSAYORQODFFVNNSKQEKADATDDTLSSKLKVKEU-SVI	688		
QY	293	EKMENOSENNDDIEVIITAPLYKQILNYAEENAIETIFYLGEALRGV---	IDLDVF	349	
Db	689	DGRRAQNCNILLSRLKSNDIEIKRALITMDEQEDLPKMI----	EQLLFVEPKSDIDL	744	
QY	350	LKHVLLSRKFQFLRALMQAR	371		
Db	745	EBHKHELRMAKAQREFLEMSR	766		
RESULT 15					
SUM2 SCHPO					
ID	SUM2 SCHPO	STANDARD;	PRT;	426 AA.	
AC	O9HGJ3;	F78821;			
DT	10-OCT-2003	(Rel. 42, Created)			
DT	10-OCT-2003	(Rel. 42, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	GM2	protein.			
GN	SUM2 OR SPEC800.09.				
OC	Schizosaccharomyces pombe (Fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OC	Schizosaccharomycetes.				
OC	NCBI_TaxId=4896;				
RP	[1]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RC	MEDLINE=21848401; PubMed=11859360;				
RA	Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,				
RA	Brooks J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,				
RA	Grout G., Brown D., Brown S., Chillingworth I., Church C.M.,				
RA	Collins M., Connor R., Cronin A., Davis P., Feldwell T., Fraser A.,				
RA	Gentiles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,				
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,				
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,				
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,				
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,				
RA	Rutherford K., Rutter S., Saunders R., Squares S., Stevens K.,				
RA	Skellton J., Simmonds M., Squares R., Squares S., Warren T., Whitehead S.,				
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Grymonprez B.,				
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,				
RA	Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,				
RA	Borzym K., Langer U., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,				
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Burnelle B.,				
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,				
RA	Galibert F., Aves S.-J., Xiang Z., Hunt C., Moore K., Hurst M.M.,				

about 49-75 repeats in DRPLA patients. Longer expansions result in earlier onset and more severe clinical manifestations of the disease.

-I- DISEASE: Defects in DRPLA are the cause of dentatorubral-pallidoluysian atrophy (DRPLA) [MIM:125370], an autosomal dominant neurodegenerative disorder characterized by a loss of neurons in the dentate nucleus, rubrum, globus pallidus and luy's/body. Clinical features are myoclonus epilepsy, dementia, and cerebellar ataxia. Onset of the disease occurs usually in the second decade of life and death in the fourth.

-I- CAUTION: Ref.2 sequence differs from that shown due to several frameshifts.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D31840; BAA06626.1; .
EMBL; D38529; BAA07534.1; ALT_FRAME.
EMBL; U23851; AAB50276.1; .
EMBL; U47924; AAB51321.1; .
EMBL; D63808; BAA23631.1; .
EMBL; L10377; -; NOT_ANNOTATED_CDS.
PIR; G01763; G01763.
Genew; HGNC:3033; DRPLA.
MIM; 607462; .
MIM; 125370; .
GO; GO:0005737; Cytoplasm; TAS.
GO; GO:0005634; Nucleus; TAS.
GO; GO:000515; P:protein binding; IPI.
GO; GO:0007417; P:central nervous system development; TAS.
InterPro; IPR002951; Atrophin.
Pfam; PF03154; Atrophin-1; 2.
PRINTS; PR01222; ATROPHIN.
Triplet repeat expansion; Polymorphism; Epilepsy.
KW SER/GLU-RICH (MIXED CHARGE).
FT DOMAIN 73 82
FT DOMAIN 302 305 POLY-PRO.
FT DOMAIN 376 382 POLY-SER.
FT DOMAIN 386 397 POLY-SER.
FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 479 483 POLY-HIS.
FT DOMAIN 484 497 POLY-GLN.
FT DOMAIN 504 507 POLY-PRO.
FT DOMAIN 564 574 POLY-SER.
FT DOMAIN 704 707 POLY-PRO.
FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).
FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).
FT CONFLICT 94 94 MISSING (IN REF. 3).
FT CONFLICT 333 333 H -> Y (IN REF. 1).
FT CONFLICT 339 339 M -> I (IN REF. 3).
FT CONFLICT 541 541 P -> T (IN REF. 6).
FT CONFLICT 1028 1028 A -> G (IN REF. 1).
FT CONFLICT 1185 AA; 124773 MM; 5135744CEE491C18 CRC64;
SEQUENCE 1185 AA; 124773 MM; 5135744CEE491C18 CRC64;
Query Match 6.3%; Score 127; DB 1; Length 1185;
Best Local Similarity 23.6%; Pred. No. 10;
Matches 64; Conservative 32; Mismatches 93; Indels 82; Gaps 13;
QY 70 YPNPPICFVKTSMITKT-----GKVDANGKI----- 99
Db 594 YFP-PPVPTV-TTSSATLSVIATVASSPAGYKTSPPGPPYKGRAPSGAYKTATPPG 651
QY 100 YLPVLHMKHPQSDLLGLIQMIVVFGDEPPVFSRPIASAPYQATGPP--NTSYMPGM 157
Db 652 YKP-----GSPFSR---TGTPPGYRGTSPPAGPTGFKGS 684
QY 158 PG-GISPPSGYPPNPSGYGCPYPP-----GGYPATNTSQYPS---QPPVTVVGSRD 209

Db 685 PTVGPGLP---PAGPSGLPSLPPPAAPASGPPLSATQIKQBPABEYETPSPVPPARS 741
QY 209 GTTSETITRASLISAVSKLRWKEEMDR-----AQELNALKRTEDLKQGHQKLEEM 263
Db 742 PSPPPKVDVPSHASQSA---RFNKHLDGRGFNSCARSDLYFVPLEGSLAKRADLIVEK 797
QY 264 VTFDLDEAVRVDKNIELLKKDELSALEK 294
Db 798 VRREAQRAEKEKEREREREKERERER 828
RESULT 17
PD61 HUMAN
ID PD61_HUMAN STANDARD; PRT; 868 AA.
AC Q8WUW4; Q9BX86; Q9NUN0; Q9P2H2; Q9UKL5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Programmed cell death 6-interacting protein (PDCD6-interacting protein) (ALG-2 interacting protein 1) (Hp95).
DE PDCD6IP OR AIP1 OR KIAA1375.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21539686; PubMed=11683497;
RA Wu Y., Pan S., Che S., He G., Nelman-Gonzalez M., Weil M.M., Kuang J.;
RT "Overexpression of Hp95 induces G1 phase arrest in confluent HeLa cells.";
RL Differentiation 67:139-153(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Li H., Shioda T., Isselbacher K.J.;
RT "Molecular cloning of human ALG-2 interacting protein 1 (AIP1).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kainine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., Labaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length CDSs in BD Creator(TM) system donor vector.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik A., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.I., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Hulse J.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002)."
RL [5]
RP SEQUENCE OF 209-868 FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

```
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Raku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.,
RA "NEDO human cDNA sequencing project.";
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL [6]
RN SEQUENCE OF 323-868 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
CC -!- FUNCTION: May play a role in the regulation of both apoptosis and
CC cell proliferation.
CC -!- SUBUNIT: Interacts with PDCD6/ALG-2 and SH3BP1 (By similarity).
CC -!- SIMILARITY: Contains 1 BRO1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use. By non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF349951; AAK20398.1; -
DR EMBL; AF151793; AAF08220.1; -
DR EMBL; BT007367; AAP36031.1; -
DR EMBL; BK020066; AAH20066.1; -
DR EMBL; AK002122; BAA92092.1; ALT_INIT.
DR EMBL; AB037796; BAA92613.1; -
DR Genew; HGNC:8766; PDCD61P.
DR MIM; 608074; -
DR InterPro; IPR004328; BRO1.
DR Pfam; PF03097; BRO1; 1.
KW Apoptosis.
FT DOMAIN 1 167
FT DOMAIN 717 860
FT CONFLICT 309 309 A -> T (IN REF. 2).
FT CONFLICT 378 378 V -> I (IN REF. 5).
FT CONFLICT 550 550 N -> S (IN REF. 1 AND 6).
FT CONFLICT 580 580 M -> T (IN REF. 5).
FT CONFLICT 730 730 S -> L (IN REF. 2).
SQ SEQUENCE 868 AA; 96023 MW; 573588D1F612EC93 CRC64;
Query Match 6.3%; Score 126; DB 1; Length 868;
Best Local Similarity 41.0%; Pred. No. 8.1;
Matches 32; Conservative 3; Mismatches 33; Indels 10; Gaps 4;
QY 139 SYPPYQATGPNTSYMPGPGIS-PYPSGYPPNPSPYPGCPYP-----GGPYRPTT 190
Db 792 SAPPPQAGPPYPT-PGYPGQCMPPMGXPNFYAYGQYNNPYPVYHQSPQAPYPGFQ 850
QY 191 SSQYP-SQPPVTTPGSR 207
Db 851 QSPYPPFPQPPQSYYPQ 868
RESULT 18
ID CALL MOUSE
AC P11087; Q06035; PRT; 1453 AA.
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COLA1.
OS Mus musculus (Mouse).
```

```
OC
OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT of type I procollagen.";
RL Matrix Biol. 14:593-595(1995).
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT collagen protein.";
RL Gene 39:311-312(1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
RT evidence for a mouse Bl element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371(1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69(1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE=88124276; PubMed=3340560;
RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT of the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773(1988).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- SIMILARITY: Contains 1 WFEC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08020; AAA88912.1; -
DR EMBL; X15896; CAA33904.1; -
DR EMBL; M14423; AAA37333.1; -
DR EMBL; M17491; AAA37334.1; -
DR EMBL; X06753; CAA29927.1; -
DR EMBL; K03036; AAA37332.1; -
DR EMBL; K03029; AAA37332.1; JOINED.
DR EMBL; K03030; AAA37332.1; JOINED.
DR EMBL; K03031; AAA37332.1; JOINED.
DR EMBL; K03032; AAA37332.1; JOINED.
DR EMBL; K03033; AAA37332.1; JOINED.
DR EMBL; K03034; AAA37332.1; JOINED.
DR EMBL; K03035; AAA37332.1; JOINED.
DR PIR; S57243; S21626
DR MGD; MGI:88467; Colla1.
DR InterPro; IPR008161; C1g_helix.
```

```

DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR002181; Fibrinogen_C.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR ProDom: PD000007; Clg_helix; 1.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWF; 1.
DR PROSITE: PS01208; VWF 1; 1.
DR PROSITE: PS0184; VWF 2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 VWF.
FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 168 1181 TRIPLE-HELICAL REGION.
FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 1450 1450 A -> V (IN REF. 5).
FT SEQUENCE 1453 AA; 137944 MW; 3B802B535DF81808 CRC64;
SQ
Query Match 6.3%; Score 126; DB 1; Length 1453;
Rest Local Similarity 24.0%; Pred. No. 15;
Matches 66; Conservative 30; Mismatches 97; Indels 82; Gaps 12;
QY 129 PVFSPRISASYPYQATGPNTSYMPMGEGISP-----YPSGYPPNPSGYPGC 178
DB 1120 FSGASGAGPRGPPGSA-GSPGKDLGLPGLPGPPGRGTGDSGPAG-PPGPPGPPG 1177
QY 179 PYPGPGYPATTSQVPSQPVTVTPGSRDGTISEDITRASLSAVSDK-----LRWR 231
DB 1178 PGPPSGGYDFSLPQPQEQ-----KSQDG---DRYRADDANVRDRLADVATLK-S 1226
QY 232 MKBEMDRAQELNALK---RTEDLKKGHQKLEMTVRLDQVAVDKNIELKKKDEEL 288
DB 1227 LSQQIENIRSPGSKNKPARTCDRLKCHSDKSGEYWD-----1266
QY 289 SSALKEMNSENNIDDEV-----IIPAPLYKQILNLYAENAIETDIFYLGEA 338
DB 1267 -----PNQGCNLDAIKVYCNMETGTCVFTPTQPSVPQ-KNMYISPNPKKKHVFGE 1318
QY 339 LRRGV-----IDLQVFLKHVLLSRKQFO 362
DB 1319 MTQGFPEYSGSDPTDVAIQLTFLRLMSTEARQ 1353
RESULT 19
PD6I RAT
ID PD6I RAT STANDARD; PRT; 401 AA.
AC Q9QZ42;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Programmed cell death 6 interacting protein (ALG-2 interacting protein
DE 1) (fragment).
GN PDCD6IP OR AIPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A. AND INTERACTION WITH SH3KBP1.
RX MEDLINE=20317131; PubMed=10858458;
RA Chen B., Borinstein S.C., Gillis J., Sykes V.W., Bogler O.;
RT "The glioma-associated protein SETA interacts with AIPI/Alx and ALG-2
and modulates apoptosis in astrocytes.";
J. Biol. Chem. 275:19275-19281(2000).
-|- FUNCTION: May play a role in the regulation of both apoptosis and
cell proliferation.
-|- SUBUNIT: Interacts with PDCD6/ALG-2 and SH3KBP1. The interaction
with PDCD6 is dependent on calcium. The activity of PDCD6-PDCD6IP
complex may be promoted by the binding of SH3KBP1 to PDCD6.
-|- TISSUE SPECIFICITY: Expressed in astrocytes and glioma cells.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF192757; AAF07179.1; -.
DR EMBL; AF192757; AAF07179.1; -.
FT NON_TER 1 393 PRO-RICH.
FT DOMAIN 250 393
SQ SEQUENCE 401 AA; 44061 MW; D9B3140782033834 CRC64;
Query Match 6.2%; Score 125; DB 1; Length 401;
Rest Local Similarity 41.0%; Pred. No. 3.8;
Matches 32; Conservative 3; Mismatches 33; Indels 10; Gaps 4;
QY 139 SYPPYQATGPNTSYMPMGEGIS-PYPSGYPPNPSGYPGCPYPP-----GGPYPAT 190
DB 325 SASPPQAQGGPPYPT-Y-PGYPGYQMPMPMGYNFYTYGQYNNMYPYVYVHQSFGQAPYGPQ 383
QY 191 SSOYP-SQPPVTVTPGSPR 207
DB 384 QTYPPFPQPPQSYFYQQ 401
RESULT 20
PD6I MOUSE
ID PD6I MOUSE STANDARD; PRT; 869 AA.
AC Q9WU78; O88695; O89014; Q8BSL8; Q8ROH5; Q99LR3; Q9QZN8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Programmed cell death 6 interacting protein (ALG-2 interacting protein
DE X) (ALG-2 interacting protein 1) (B2F1-inducible protein) (Eig2).
GN PDCD6IP OR ALIX OR AIPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND INTERACTION WITH PDCD6.
TISSUE=Brain;
RX MEDLINE=99218669; PubMed=10200558;
RA Missotten M., Nichols A., Rieger K., Sadoul R.;
RT "Alx, a novel mouse protein undergoing calcium-dependent interaction
with the apoptotic-linked-gene (ALG-2) protein.";
Cell Death Differ. 6:124-129(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=99098896; PubMed=9880530;
RA Vito P., Pellegrini L., Guet C., D'Adamo L.;
RT "Cloning of AIPI, a novel protein that associates with the apoptosis-
linked gene ALG-2 in a Ca2+-dependent reaction.";
J. Biol. Chem. 274:1533-1540(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RX STRAIN=C57BL/6J; TISSUE=Forelimb;
RA MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,

```

RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Perteu G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Vetraro R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wu Ming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RP SEQUENCE OF 671-869 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Forelimb;
RX MEDLINE=20027348; PubMed=10556317;
RA Wang A.J., Pierce A., Judson-Kremer K., Gaddis S., Aldaz C.M.,
RA Johnson D.G., MacLeod M.C.;
RT "Rapid analysis of gene expression (RAGE) facilitates universal
RT expression profiling.";
RL Nucleic Acids Res. 27:4609-4618 (1999).
CC -!- FUNCTION: May play a role in the regulation of both apoptosis and
CC cell proliferation.
CC -!- SUBUNIT: Interacts with PDCD6/ALG-2 and SH3BP1 (By similarity).
CC The interaction with PDCD6 is dependent on calcium.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9UW78-1; Sequence=Displayed;
CC Name=2; Synonyms=Alix-SF, Short;
CC IsoId=Q9UW78-2; Sequence=VSP_007502;
CC Note=Does not interact with ALG-2;
CC Name=3;
CC IsoId=Q9UW78-3; Sequence=VSP_007501;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -!- SIMILARITY: Contains 1 BR01 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005073; CAA06329.1; -
DR EMBL; AJ005074; CAA06330.1; -
DR EMBL; AF119955; AAD26813.1; -
DR EMBL; AK031256; BAC27323.1; -
DR EMBL; BC002261; AAH02261.1; -
DR EMBL; BC026823; AAH26823.1; -
DR EMBL; AF176514; AAD53115.1; -
DR MGD; MGI-1333753; Pdc6p.1; -
DR GO; GO:0005829; C:cytosol; IDA.
DR InterPro; IPR004328; BR01.
DR Pfam; PF03097; BR01; 1.
KW Apoptosis; Alternative splicing.
FT DOMAIN 1 167
FT PRO-RICH
FT K -> KYFYQ (in isoform 3).
FT VARSPPLIC 239 239
FT Missing (in isoform 2).
FT /FTid=VSP_007501.
FT /FTid=VSP_007502.
FT LDPIG -> SGYSR (IN REF. 2).
FT V -> L (IN REF. 2, 3 AND 4).
FT EV -> DL (IN REF. 2).
FT N -> T (IN REF. 4).
FT L -> V (IN REF. 2).
FT KQ -> NE (IN REF. 2).
FT G -> R (IN REF. 2).
FT P -> L (IN REF. 1; CAA06330).
SQ SEQUENCE 869 AA; 96010 MW; 9AA84B592FDC9E CRC64;
Query Match 6.2%; Score 125; DB 1; Length 869;
Best Local Similarity 41.0%; Pred. No. 9.2;
Matches 32; Conservative 3; Mismatches 33; Indels 10; Gaps 4;
QY 139 SYPPYQATGPTNSYMPGMPGSGIS-PYPSGYPPPSGYPGCPYPP-----GGYPYPTT 190
Db 793 SAPPPQAGGPYPTT-PGYPGYCOMPMPGMYNPAYGQNNMYPVYHQSGQAFYPGFQ 851
QY 191 SSQYP-SQPPVTTVGPSR 207
Db 852 QTYVFPQPPQSQSYYPQ 869
RESULT 21
OSA DROME
ID OSA DROME STANDARD; PRT; 2716 AA.
AC Q81N94; Q61603; Q9VEG7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trithorax group protein OSA (Eyelid protein).
GN OSA OR ELD OR CG7467.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.; FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=97415319; PubMed=9271118;
RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RT "eyelid antagonizes wingless signaling during Drosophila development
RT and has homology to the Bright family of DNA-binding proteins.";
RL Genes Dev. 11:1949-1962 (1997).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.R., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Berson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkian I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [3]
RN FUNCTION, AND DEVELOPMENTAL STAGE.
RX MEDLINE=99112962; PubMed=9895321;
RA Vazquez M., Moore L., Kennison J.A.;
RT "The trithorax group gene osa encodes an ARID-domain protein that
RT genetically interacts with the Brahma chromatin-remodeling factor to
RT regulate transcription.";
RL Development 126:733-742 (1999).
RN [4]
RN DNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.
RX MEDLINE=20069333; PubMed=10601025;
RA Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
RT "Osa associates with the Brahma chromatin remodeling complex and
RT promotes the activation of some target genes.";
RL EMBO J. 18:7029-7040 (1999).
RN [5]
RN FUNCTION
RX MEDLINE=99403006; PubMed=10471712;
RA Staehling-Hampton K., Ciampa P.G., Brook A., Dyson N.;
RT "A genetic screen for modifiers of E2f in Drosophila melanogaster.";
RL Genetics 153:275-287 (1999).
RN [6]
RN IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM; OSA; MOR; SNR1; DALAO;
RP BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
RX MEDLINE=20270023; PubMed=10809685;
RA Kai A.J., Mahmoudi T., Zak N.B., Verrilizer C.P.;
RT "The Drosophila brahma complex is an essential coactivator for the
RT trithorax group protein zeste.";
RL Genes Dev. 14:1058-1071 (2000).
RN [7]
RN FUNCTION AS A COREPRESSOR.
RP MEDLINE=20573925; PubMed=11124806;
RX

RA Collins R.T., Treisman J.E.;
RT "Osa-containing Brahma chromatin remodeling complexes are required
RT for the repression of wingless target genes.";
RL Genes Dev. 14:3140-3152 (2000).
RN [8]
RN FUNCTION AS A COREPRESSOR, AND INTERACTION WITH PNR AND CHI.
RX MEDLINE=22515897; PubMed=12629041;
RA Heitzler P., Vanolst L., Biryukova I., Raman P.;
RT "Enhancer-promoter communication mediated by Chip during
RT Pannier-driven proneural patterning is regulated by Osa.";
RL Genes Dev. 17:591-596 (2003).
CC -1- FUNCTION: Trithorax group (trxG) protein required for embryonic
CC segmentation, development of the notum and wing margin, and
CC photoreceptor differentiation. Required for the activation of
CC genes such as Anp, Ubx and Eve. Binds to DNA without specific
CC affinity, suggesting that it is recruited to promoters by
CC promoter-specific proteins. Essential component of the Brahma
CC complex, a multiprotein complex which is the equivalent of the
CC yeast SWI/SNF complex and acts by remodelling the chromatin by
CC catalyzing an ATP-dependent alteration in the structure of
CC nucleosomal DNA. This complex can both serve as a transcriptional
CC coactivator or corepressor, depending on the context. Acts as an
CC essential coactivator for Zeste, which recruits the whole complex
CC to specific genes. In contrast, it acts as a corepressor for Wg
CC target genes, possibly via an interaction with Pan and Gro. It
CC also acts as a negative regulator for proneural achaete-scute,
CC when it is directly recruited by Pan and Chi. Also represses E2f
CC activation.
CC -1- SUBUNIT: Component of the Brahma complex, which is composed of
CC Brm, Osa, Mor, Snr1/Bap45, Bap11/Dalao, Bap55, Bap60 and Bap47.
CC Interacts with Pnr and Chi via its EHD domain.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in early embryo. In
CC third instar larvae, it is ubiquitously expressed in wing and eye-
CC antenna imaginal disks, with a stronger expression in a band just
CC anterior to the morphogenetic furrow.
CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC -1- DOMAIN: The ARID domains mediates the binding to DNA.
CC -1- SIMILARITY: Contains 1 ARID domain.
CC -1- SIMILARITY: Contains 1 EHD (Eyelid homology) domain.
CC -1- CAUTION: Ref.2 (AAF55457) sequence differs from that shown due to
CC erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL; AF053091; AAC06254.1; -;
CC EMBL; AS003718; AAF55457.1; ALT_SEQ.
CC EMBL; AS003718; AAN13750.1; -;
CC PIR; T13049; T13049.
CC FlyBase; FBgn0003013; osa.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0003677; F:DNA binding; IDA.
CC GO; GO:004530; P:photoreceptor cell differentiation; IMP.
CC GO; GO:0045449; P:regulation of transcription; IDA.
CC GO; GO:0007379; P:segment specification; IMP.
CC GO; GO:0008587; P:wing margin morphogenesis; IMP.
CC GO; GO:0016055; P:wnt receptor signaling pathway; IMP.
CC InterPro; IPR001606; ARID.
CC InterPro; IPR008938; ARM.
CC InterPro; IPR006031; XYPPX.
CC Pfam; PF01388; ARID; 1.
CC Pfam; PF02162; XYPPX; 8.
CC SMART; SM00501; BRIGHT; 1.
CC Transcription regulation; DNA-binding; Activator; Repressor;
CC Chromatin regulator; Nuclear protein; Developmental protein.
CC DOMAIN 997 1111 ARID.
CC DOMAIN 1769 2517 EHD.
CC FT


```
FT DOMAIN 19 1763 PRO-RICH.
FT DOMAIN 174 380 GIN-RICH.
FT DOMAIN 619 873 GLY-RICH.
FT DOMAIN 1222 1453 GLY-RICH.
FT DOMAIN 1271 1751 HIS-RICH.
FT DOMAIN 1730 1745 HIS-RICH.
FT DOMAIN 2589 2624 SER-RICH.
FT DOMAIN 2625 2716 ALA-RICH.
FT CONFLICT 61 61 MISSING (IN REF. 1).
FT CONFLICT 1169 1169 V -> G (IN REF. 1).
FT CONFLICT 1795 1795 M -> T (IN REF. 1).
FT CONFLICT 2637 2637 G -> E (IN REF. 1).
SQ SEQUENCE 2716 AA; 284063 MW; 8FAE76CB51C7C675 CRC64;

Query Match 6.2%; Score 125; DB 1; Length 2716;
Best Local Similarity 30.3%; Pred. No. 34;
Matches 43; Conservative 8; Mismatches 43; Indels 48; Gaps 8;

QY 71 PYNPICFVKPTSSMTIKTKGKVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVFGDEPPP 130
Db 1258 PHPPP-----PHSPHTAAQ-----QAAGQ-----HQQHPQHPEGL-----PGPP 1293
QY 131 VSRPISASYPYQATGPTNTSYMPGPGGIGSPYSGY-----PNPNSGY-----PGCPYP 181
Db 1294 -----PPQQOQGGQGPFPVSGGPPPAQHQHGGVPPSPQOHVRPAAGAYP 1343
QY 182 PGG-----PYPATTSSQYPSQP 198
Db 1344 PGGSGYPTPVSRTPGSPYPSQP 1365

RESULT 22
YQYL CAEEL STANDARD; PRT; 964 AA.
AC Q09560;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 108.9 kDa protein F36G3.1 in chromosome X.
GN F36G3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RA White S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RA Jones S.J.M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 247069; CAA87338.1; -.
DR F01; T21865; T21865.
DR WormPep; F36G3.1; CEI5979.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 656 840 COILED COIL (POTENTIAL).
SQ SEQUENCE 964 AA; 108868 MW; D8A6BC59350F076 CRC64;

Query Match 6.2%; Score 124; DB 1; Length 964;
Best Local Similarity 25.3%; Pred. No. 12;
Matches 50; Conservative 46; Mismatches 46; Indels 36; Gaps 9;

QY 207 RDGTISDITRASLISAVSDKLRWRMKEMDMRAQALNALKTEEDLKKGHQKLEEMVTR 266
Db 636 QDQTQTDNNNSFILVDHRHVPVMSLQVEIDRLQDLEKVKSGKELLQINSKLEELDE 695
QY 267 LDOEV-----AEVDKNIELLKKKDELSALEKMN-----QSENNIDID- 305
Db 696 SQQTDIGIEIEAEQOYTSLTSEIDELCHIVRQDQELAILKEKVTNVININSLDKDDVDS 755
QY 306 -EVIITAPLYKQIL-NLYAENAJEDTIFYLGEA--LRRGVLDLDVFLKHVRLLSRKQF 361
Db 756 QXVIVQR---QKEIENREELDAITKK---LGEVTKLRDKAVE--EATLYKMKMWERDRF 808
QY 362 QLRALMQKARKTAGLSDL 379
Db 809 ----LSREAQMSMEIEDL 822

RESULT 23
VASP HUMAN STANDARD; PRT; 379 AA.
AC P50552; Q93035;
DT 01-OCT-1996 (Rel. 34, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vasodilator-stimulated phosphoprotein (VASP).
GN VASP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95129547; PubMed=7828592;
RA Haffner C., Jarchau T., Reinhard M., Hoppe J., Lohmann S.M.,
RA Walter U.;
RT "Molecular cloning, structural analysis and functional expression of
RT the proline-rich focal adhesion and microfilament-associated protein
RT VASP."
RL EMBO J. 14:19-27(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung, and Fetal spleen;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards D., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketchum M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 2-379 FROM N.A.
RX MEDLINE=96411679; PubMed=8812448;
RA Zimmer M., Fink T., Fischer L., Hauser W., Scherer K., Lichter P.,
RA Walter U.;
RT "Cloning of the VASP (vasodilator-stimulated phosphoprotein) genes in
RT human and mouse: structure, sequence, and chromosomal localization."
RL Genomics 36:227-233(1996).
RN [4]
RP SEQUENCE OF 1-9.
```



```

RN Science 287:2185-2195 (2000).
RP [2]
RC SEQUENCE OF 980-2779 FROM N.A.
RA STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Quarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RN Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
RP [3]
RP FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20530668; PubMed=11076973;
RA Sisson J.C., Field C., Ventura R., Royou A., Sullivan W.;
RT "Lava lamp, a novel peripheral golgi protein, is required for
RT Drosophila melanogaster cellularization.";
RN J. Cell Biol. 151:905-918 (2000).
CC -!- FUNCTION: Lva and spectrin may form a Golgi-based scaffold that
CC mediates interaction of Golgi bodies with microtubules and
CC facilitates Golgi-derived membrane secretion required for the
CC formation of furrows during cellularization.
CC -!- SUBUNIT: Interacts with CLIP-190 and spectrin separately.
CC -!- SUBCELLULAR LOCATION: Lva-alpha-spectrin and Lva-CLIP-190
CC complexes are found at the Golgi.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003429; AAF45910.1; -.
DR EMBL; AY118638; AAM50007.1; ALT_INIT.
DR FlyBase; FBgn0029688; lva.
KW Developmental protein; Golgi stack; Coiled coil.
FT DOMAIN 52 85 COILED COIL (POTENTIAL).
FT DOMAIN 141 175 COILED COIL (POTENTIAL).
FT DOMAIN 220 607 COILED COIL (POTENTIAL).
FT DOMAIN 659 716 COILED COIL (POTENTIAL).
FT DOMAIN 751 1733 COILED COIL (POTENTIAL).
FT DOMAIN 1785 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1941 2433 COILED COIL (POTENTIAL).
FT DOMAIN 2504 2544 COILED COIL (POTENTIAL).
FT DOMAIN 2600 2641 COILED COIL (POTENTIAL).
FT DOMAIN 1211 1211 E -> K (IN REF. 2).
FT CONFLICT 1424 1424 A -> S (IN REF. 2).
FT CONFLICT 1506 1506 E -> D (IN REF. 2).
FT CONFLICT 2034 2034 A -> V (IN REF. 2).
FT CONFLICT 2069 2069 Q -> H (IN REF. 2).
FT CONFLICT 2133 2133 Q -> E (IN REF. 2).
FT CONFLICT 2154 2154 Q -> V (IN REF. 2).
FT CONFLICT 2160 2160 L -> P (IN REF. 2).
FT CONFLICT 2200 2200 E -> T (IN REF. 2).
FT CONFLICT 2217 2217 E -> D (IN REF. 2).
FT CONFLICT 2271 2271
SQ SEQUENCE 2779 AA; 315897 MW; 1CB3965102018ABE CRC64;

Query Match
Best Local Similarity 6.2%; Score 123.5; DB 1; Length 2779;
Matches 48; Conservative 33; Mismatches 58; Indels 39; Gaps 8;
QY 199 PVTTPGSRDGTISDTIRASLISAVSKLRWRMKE---EMDRAQAEINLAKRTTE--DL 253
DQ 2088 PAATVSSSLDG---DEAARIESLQEIQLRQCVSELEDERTRDQAEIQAALRQSSQGYDE 2144
QY 254 KGHQKLEFEMVTRLDQAEVAVDVKNEILLKKDEBELSSALKRMENSENNDDEVIIPTAP 313
DQ 2145 AEDNQKLEL-----QQLRQSELEALRTRDQ---SELEALRQSCQGH-ETVRIATLQ 2194
QY 314 LVKQLNLVAEENATEDITFYLGELALRGVIDLDVFLKRVLLSRKQFQLRALMQKAR 371
-----
Db 2195 QNQOQLEL-----QQLRQAIIETLET-----LRARDQTELEALROSSQ 2231
RESULT 25
DIAL HUMAN
ID DIAL HUMAN STANDARD; PRT; 1248 AA.
AC C60610; Q9UC76;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Diaphanous protein homolog 1 (Diaphanous-related formin 1) (DRF1).
GN DIAPH1 OR DIAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A. PubMed=9360932;
RX MEDLINE=98028756;
RA Lynch E.D., Lee M.K., Morrow J.E., Welcsh P.L., Leon P.E., King M.-C.;
RT "Nonsyndromic deafness DFNA1 associated with mutation of a human
RT homolog of the Drosophila gene diaphanous.";
RL Science 278:1315-1318 (1997).
RN [2]
RP SEQUENCE OF 218-817 FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDD human cDNA sequencing project.";
RN [3]
RP Submitted (AUG-2000) to the EMBL/GenBank/DDBB databases.
RP SEQUENCE OF 727-765 AND 1121-1145.
RC TISSUE=Platelet;
RX MEDLINE=95255215; PubMed=7737110;
RA Reinhard M., Giehl K., Abel K., Haffner C., Jarchau T., Hoppe V.,
RA Jockusch B.M., Walter U.;
RT "The proline-rich focal adhesion and microfilament protein VASP is a
RT ligand for profilins.";
RL EMBL J. 14:1583-1589 (1995).
CC -!- FUNCTION: Binds to GTP-bound form of Rho and to profilin. Acts in
CC a Rho-dependent manner to recruit profilin to the membrane, where
CC it promotes actin polymerization. It is required for cytokinesis,
CC stress fiber formation, and transcriptional activation of the
CC serum response factor. DRF proteins couple Rho and Src tyrosine
CC kinase during signaling and the regulation of actin dynamics (By
CC similarity). In hearing it may play a role in the regulation of
CC actin polymerization in hair cells.
CC -!- TISSUE SPECIFICITY: Expressed in brain, heart, placenta, lung,
CC kidney, pancreas, liver, skeletal muscle and cochlea.
CC -!- DOMAIN: DRFs are regulated by intramolecular GBD-DAD binding where
CC Rho-GTP activates the DRFs by disrupting the GBD-DAD interaction
CC (By similarity).
CC -!- DISEASE: Defects in DIAPH1 are a cause of autosomal dominant
CC nonsyndromic sensorineural deafness 1 (DFNA1) [MIM:124900].
CC -!- SIMILARITY: Contains 1 GTPase-binding (GBD) domain.
CC -!- SIMILARITY: Contains 1 Formin homology 1 (FH1) domain.
CC -!- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.
CC -!- SIMILARITY: Contains 1 Formin homology 3 (FH3) domain.
CC -!- SIMILARITY: Contains 1 DRF autoregulatory (DAD) domain.
CC -!- SIMILARITY: Belongs to the formin homology family. Diaphanous
CC subfamily.
CC -!- CAUTION: Ref.2 sequence differs from that shown in that it seems
CC to include intronic sequence.
CC -!- DATABASE: NAME=Hereditary hearing loss homepage;
CC NOTE=Gene page;
CC WWW="http://www.uia.ac.be/dnalab/hhh/".
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 50.5664 Seconds
(without alignments)
2128.898 Million cell updates/sec

Title: US-09-804-690-2

Perfect score: 1996

Sequence: 1 MMSKYYKRLTVRQTNVIA.....FQLRALMQKARKTAGLSLDLY 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A Geneseq_29Jan04.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1996	100.0	381	2	AAW19110 Mouse tum
2	1996	100.0	381	2	AAW93425 Mouse tsq
3	1996	100.0	381	6	ABG73788 Murine ts
4	1996	100.0	381	7	ADC35178 Human tsq
5	1996	100.0	391	6	AAE34884 Mouse Tsg
6	1996	100.0	391	7	ADB85228 Mouse tum
7	1900.5	95.2	380	2	AAW19111 Human tum
8	1900.5	95.2	390	2	AAW93424 Human NHT
9	1900.5	95.2	390	5	ABP52191 Human tum
10	1900.5	95.2	390	6	ABG73787 Human tum
11	1900.5	95.2	390	7	ADC35176 Novel hum
12	1895.5	95.0	390	6	AAE34885 Human Tsg
13	1337.5	67.0	307	5	ABP41729 Human ova
14	1022.5	51.2	237	4	AAE09328 Human int
15	786	39.4	331	4	ABM64607 Drosophil
16	440	22.0	398	3	AAAG06370 Arabidops
17	440	22.0	412	3	AAAG06369 Arabidops
18	415	20.8	379	5	ABP43551 L-lactate
19	406.5	20.4	340	3	AAAG06371 Arabidops
20	402	20.1	90	3	AAAG01689 Human sec
21	364.5	18.3	322	3	AAAG13898 Arabidops
22	361	18.1	341	4	AAAB93473 Human pro
23	327	16.4	146	3	AAAB53717 Human col
24	318	15.9	452	4	AAU23116 Novel hum
25	266	13.3	288	3	AAAG13899 Arabidops

26	162.5	8.1	168	3	AAAG13900 Arabidops
27	152	7.6	218	4	ABG18833 Novel hum
28	152	7.6	485	3	AAAB57086 Human pro
29	150	7.5	466	6	ABO07215 Human p53
30	150	7.5	466	7	ADD47589 Human pro
31	146	7.3	148	3	AAAY86515 Human gen
32	146	7.3	148	5	ABB97431 Novel hum
33	146	7.3	148	6	ABO53679 Novel hum
34	146	7.3	176	4	AAAG75132 Human col
35	143	7.2	777	5	ABP41839 Human ova
36	143	7.2	1130	7	ADC99065 Human KPP
37	142.5	7.1	905	2	AAW31186 Human p16
38	142.5	7.1	1135	2	AAW31185 Human p16
39	142	7.1	466	2	AAAR07084 Recombina
40	138	6.9	258	3	AAAB44500 Plant vir
41	137.5	6.9	244	3	AAAG40015 Arabidops
42	137.5	6.9	245	3	AAAG38059 Arabidops
43	137.5	6.9	245	3	AAAG05889 Arabidops
44	137.5	6.9	247	3	AAAG38058 Arabidops
45	137.5	6.9	263	3	AAAG38057 Arabidops
46	137.5	6.9	274	3	AAAG05887 Arabidops
47	137.5	6.9	1251	7	ADE09332 Novel pro
48	137.5	6.9	397	4	AAAM39852 Human pol
49	137	6.9	397	4	AAAG81381 Human APP
50	137	6.9	397	6	ADA55415 Human pro
51	137	6.9	442	4	AAAM41638 Human pol
52	137	6.9	463	7	ADD47587 Rat Prote
53	135.5	6.8	463	7	ABP35602 Fungal ZB
54	135.5	6.8	827	5	ABP35602 Fungal ZB
55	135	6.8	172	4	ABG18832 Novel hum
56	133	6.7	505	2	AAW53808 N-Wiskott
57	133	6.7	505	4	AAAG67360 Amino aci
58	133	6.7	505	4	AAAM52320 Bovine N-
59	130	6.5	787	7	ADB65186 Human pro
60	128.5	6.4	114	5	ABG77261 Selected
61	128.5	6.4	177	5	ABJ11132 Yeast sel
62	128	6.4	505	2	AAAY07117 Lung canc
63	128	6.4	505	6	ABO07216 Human p53
64	126.5	6.3	416	5	ABG70108 Human pre
65	126.5	6.3	505	2	AAW46889 Human Neu
66	126.5	6.3	505	4	AAAG67338 Amino aci
67	126.5	6.3	505	4	AAAM52316 Human N-W
68	126.5	6.3	505	6	AAAE38438 Human N-W
69	126.5	6.3	582	4	ABG16505 Novel hum
70	126	6.3	104	5	ABB04515 Murine ap
71	126	6.3	901	4	AAAM41217 Human pol
72	125.5	6.3	1429	4	ABB69606 Drosophil
73	124.5	6.2	369	7	ADD14178 Human src
74	124.5	6.2	371	2	AAW64535 Human leu
75	124.5	6.2	371	5	ABP61444 Human NF-
76	124.5	6.2	371	6	ABR41017 Human MAP
77	123.5	6.2	1040	3	AAAG30935 Arabidops
78	123.5	6.2	1052	3	AAAG30934 Arabidops
79	123.5	6.2	1181	3	AAAG30933 Arabidops
80	123	6.2	744	6	ABU34940 Proteir e
81	122.5	6.1	501	2	AAW46890 Rat Neura
82	122.5	6.1	501	4	AAAG67355 Amino aci
83	122.5	6.1	501	4	AAAM52319 Rat N-WAS
84	122.5	6.1	501	6	AAAE38439 Rat N-WAS
85	122.5	6.1	503	2	AAAR34127 Annexin X
86	122.5	6.1	505	2	AAAR34128 Annexin X
87	122.5	6.1	1042	7	ADB80342 Human MDD
88	122.5	6.1	1141	6	ABR39812 Human SCA
89	122.5	6.1	1196	6	ABR39811 Human SCA
90	122	6.1	171	2	AAR27476 20 repeat
91	122	6.1	420	5	ABG93946 Dear poly
92	122	6.1	748	4	AAAM40299 Human pol
93	122	6.1	1104	7	AAAE37912 Human CGD
94	121.5	6.1	214	2	AAAR6913 Cotton fi
95	121	6.1	612	4	AAAB95546 Human pro
96	121	6.1	887	7	AADE55580 Human pro
97	121	6.1	1463	4	AAE02532 Bovine al
98	121	6.1	1617	3	AAAG52174 Arabidops

99 121 6.1 1623 3 AAG52173 Arabidops
100 121 6.1 1683 3 AAG52172 Arabidops

ALIGNMENTS

```

RESULT 1
AAW19110
ID AAW19110 standard; protein; 381 AA.
XX
AC AAW19110;
XX
DT 10-DEC-1997 (first entry)
XX
DE Mouse tumour susceptibility protein TSG101.
XX
KW Tumour susceptibility protein; TSG101; tumorigenesis; breast cancer;
diagnosis; therapy; mouse.
XX
OS Mus musculus.
XX
FH Key
FT Modified-site 11 Location/Qualifiers
FT Region /note= "potential protein kinase C phosphorylation site"
FT /note= "residues 37-46 resemble a helix-turn-helix
signature domain"
FT Modified-site 38
FT Modified-site 38 /note= "potential casein kinase II phosphorylation site"
FT Modified-site 44 /note= "potential protein kinase C phosphorylation site"
FT Modified-site 55 /note= "potential N-glycosylation site"
FT Region /note= "potential N-myristoylation site"
FT /note= "residues 73-83 resemble a fungal Zn-cys bi-
nuclear cluster signature "
FT Modified-site 85
FT Modified-site 88 /note= "potential protein kinase C phosphorylation site"
FT Modified-site 150 /note= "potential protein kinase C phosphorylation site"
FT Modified-site 156 /note= "potential N-glycosylation site"
FT Modified-site 210 /note= "potential N-myristoylation site"
FT Modified-site 215 /note= "potential casein kinase II phosphorylation site"
FT Modified-site 225 /note= "potential protein kinase C phosphorylation site"
FT Modified-site 249 /note= "potential protein kinase C phosphorylation site"
FT Modified-site 265 /note= "potential casein kinase II phosphorylation site"
FT Modified-site 290 /note= "potential casein kinase II phosphorylation site"
FT Modified-site 297 /note= "potential casein kinase II phosphorylation site"
FT Modified-site 357 /note= "potential N-glycosylation site"
FT /note= "potential protein kinase C phosphorylation site"
XX
PN W09718333-A1.
XX
PD 22-MAY-1997.
XX
PF 15-NOV-1996; 96WO-US018828.
XX
PR 16-NOV-1995; 95US-0006856P.
16-JAN-1996; 96US-00585758.
13-JUN-1996; 96US-00670274.

```

```

XX (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
XX Li L, Cohen SN;
XX
XX WPI; 1997-289308/26.
XX N-PSDB; AAT69574.
XX
XX Nucleic acid encoding tumour susceptibility protein TSG101 - useful to
develop products for cancer detection, prophylaxis, therapy and
predisposition determination.
XX
XX Example 1; Page 67-70; 89pp; English.
XX
XX This polypeptide sequence comprises a mouse tumour susceptibility
protein, designated TSG101, that can function as a tumour suppressor in
oncogenesis. Its amino acid sequence was deduced from a cDNA clone
(AAT69574) isolated using a novel random homomologous knockout approach. It
shows 94% identity to human TSG101 (see AAW19111). Deletions in human
TSG101 are associated with the occurrence of cancers, e.g. breast cancer.
Methods are provided for identifying homologous or related proteins, for
producing compositions that modulate the expression or function of the
TSG101 protein, and for studying physiological pathways. Modulation of
gene activity in vivo is used for prophylactic or therapeutic purposes,
such as cancer therapy, identification of cell type based on expression,
etc
XX
SQ Sequence 381 AA;
Query Match 100.0%; Score 1996; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.3e-145;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMSKYYRDLTVRQTVNVVIAMKYDLKPVLDSSYVNDGSSRELNVLTGTIPVYRGNTYNI 60
DB 1 MMSKYYRDLTVRQTVNVVIAMKYDLKPVLDSSYVNDGSSRELNVLTGTIPVYRGNTYNI 60
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDWKHPRSELLELIQI 120
DB 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDWKHPRSELLELIQI 120
QY 121 MIVIFGEPPVFSRPTVSASYPPTATGPNTSYMPGMSGISAYSPGYPNPSGYPGCP 180
DB 121 MIVIFGEPPVFSRPTVSASYPPTATGPNTSYMPGMSGISAYSPGYPNPSGYPGCP 180
QY 181 YPPAGYPATTSSQYPSQPPVTVGPSRDGTISDITRASLIISAVSDKLRWRKEMDGA 240
DB 181 YPPAGYPATTSSQYPSQPPVTVGPSRDGTISDITRASLIISAVSDKLRWRKEMDGA 240
QY 241 QAEINALKRTEEDLKGHKLEEMVTRLDQEVADVDKNIELLKKDDELSALEKMNQS 300
DB 241 QAEINALKRTEEDLKGHKLEEMVTRLDQEVADVDKNIELLKKDDELSALEKMNQS 300
QY 301 ENNDIDEVIPTAPLYKQILNLYAENNAIEDTTFYLGELARRGVLDLDFLKHVRLLSRK 360
DB 301 ENNDIDEVIPTAPLYKQILNLYAENNAIEDTTFYLGELARRGVLDLDFLKHVRLLSRK 360
QY 361 OFQIRALMOKARKTAGLSDIY 381
DB 361 OFQIRALMOKARKTAGLSDIY 381
RESULT 2
AAW93425
ID AAW93425 standard; protein; 381 AA.
XX
XX AAW93425;
XX
XX 11-JUN-1999 (first entry)
XX
XX Mouse tsgl01 protein.
XX
XX Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
KW

```

KW brain; breast; colon; heart; kidney; ovary; paragonia; pancreas;
 KW prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
 KW biliary cirrhosis; Crohn's disease; rheumatoid arthritis; mouse; ts101.

XX Mus sp.

XX US5892016-A.

XX 06-APR-1999.

XX 23-JAN-1997; 97US-00786999.

XX 23-JAN-1997; 97US-00786999.

XX (INCY-) INCYTE PHARM.

XX Brie SL, Goli SK;

XX WPI; 1999-253932/21.

XX Novel human tumor suppressor - useful for the diagnosis or treatment of
 PT lymphoma, cancer, and autoimmune disease.

XX Disclosure; Fig 2; 25pp; English.

XX This invention describes the isolation of a novel human tumor suppressor
 CC (NHTS). The products of the invention may be used for the diagnosis or
 CC treatment of conditions and diseases which are associated with expression
 CC of NHTS e.g. lymphoma and cancers of the brain, breast, colon, heart,
 CC kidney, ovary, paragonia, pancreas, prostate, skin, stomach and thyroid
 CC and autoimmune disease e.g. asthma, biliary cirrhosis, Crohn's disease,
 CC diabetes, and rheumatoid arthritis

XX Sequence 381 AA;

Query Match 100.0%; Score 1996; DB 2; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.3e-145;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMSKYKRDLTVRQTVNVIAMKYDKLPVLDYVFNDSGSSRELNLVTGTIPVYRGNIYNI 60
 Db 1 MMSKYKRDLTVRQTVNVIAMKYDKLPVLDYVFNDSGSSRELNLVTGTIPVYRGNIYNI 60

QY 61 PICLWLLDTYPNPPICFVKPTSSMTIKTKHVDANGKIYLPYLDHDKHPRSELLELIQI 120
 Db 61 PICLWLLDTYPNPPICFVKPTSSMTIKTKHVDANGKIYLPYLDHDKHPRSELLELIQI 120

QY 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180
 Db 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180

QY 181 YPPAGPYPATSSQYPSQPPVTTVGPSSRDGTISEDTIRASLIISAVSDKLWRMKEMDGA 240
 Db 181 YPPAGPYPATSSQYPSQPPVTTVGPSSRDGTISEDTIRASLIISAVSDKLWRMKEMDGA 240

QY 241 QAEINLAKRTEDLKKHOKLEEMVTRLDQAEVVDKNIKLLKKDELSALEKMNQS 300
 Db 241 QAEINLAKRTEDLKKHOKLEEMVTRLDQAEVVDKNIKLLKKDELSALEKMNQS 300

QY 301 ENNDIDEVIIPTAPLYKQILNLYAEENAIETIFVILGEALRRGVTDLDVFLKHVRLLSRK 360
 Db 301 ENNDIDEVIIPTAPLYKQILNLYAEENAIETIFVILGEALRRGVTDLDVFLKHVRLLSRK 360

QY 361 QFQRLMOKARKTAGLSLDLY 381
 Db 361 QFQRLMOKARKTAGLSLDLY 381

RESULT 3
 ABG73788
 ID ABG73788 standard; protein; 381 AA.
 XX
 AC ABG73788;

XX DT

10-APR-2003 (first entry)

XX DE

Murine ts101 protein.

XX KW

Murine; tumor suppressor; NHTS; cytostatic; antiasthmatic; hepatotropic;
 KW antiinflammatory; antidiabetic; antirheumatic; antiarthritic; cancer;
 KW autoimmune disease; asthma; biliary cirrhosis; Crohn's disease; diabetes;
 KW rheumatoid arthritis; drug screening; vaccine; gene therapy; ts101.

XX OS

Mus musculus.

XX PN

US6472508-B1.

XX PD

29-OCT-2002.

XX PF

18-DEC-1998; 98US-00216387.

XX PR

23-JAN-1997; 97US-00786999.

XX PA

(INCY-) INCYTE GENOMICS INC.

XX PI

La Brie S, Goli SK;

XX WPI; 2003-196850/19.

XX New purified human tumor suppressor polypeptide, useful for diagnosing,
 PT preventing and treating cancer and autoimmune diseases (e.g. asthma or
 PT diabetes), and in drug screening assays.

XX PS

Disclosure; Fig 2; 26pp; English.

XX CC

This invention describes a novel human tumor suppressor, NHTS, which has
 CC cytostatic, antiasthmatic, hepatotropic, antiinflammatory, antidiabetic,
 CC antirheumatic and antiarthritic activity. The NHTS polypeptide is useful
 CC in diagnosing, preventing and treating cancer and autoimmune diseases
 CC (e.g. asthma, biliary cirrhosis, Crohn's disease, diabetes or rheumatoid
 CC arthritis), in drug screening assays, in vaccines and for gene therapy.
 CC The nucleic acid encoding the above polypeptide may be used to detect and
 CC quantitate gene expression in biopsied tissues in which expression of the
 CC polypeptide may be correlated with disease, and thus, may be used to
 CC distinguish between absence, presence, and excess expression of the
 CC polypeptide, and to monitor the regulation of the polypeptide levels
 CC during therapeutic intervention. The polynucleotide may also be used to
 CC generate hybridisation probes for mapping genomic sequences, and in
 CC various molecular biology techniques. Antibodies generated from the
 CC products of the invention may be used as a targeting or delivery
 CC mechanism for bringing an agonist or other pharmaceutical agent to cells
 CC of patients having the above-mentioned diseases, and in diagnosing
 CC conditions or diseases characterised by the expression of the
 CC polypeptide, or in assays to monitor patients being treated with the
 CC above polypeptide, agonists or antagonists. This sequence represents the
 CC murine ts101 protein, a homologue of the human NHTS protein, isolated
 CC from Incyte clone 609476 derived from COLN0T01, a colon cDNA library

XX SQ

Sequence 381 AA;

Query Match 100.0%; Score 1996; DB 6; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.3e-145;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMSKYKRDLTVRQTVNVIAMKYDKLPVLDYVFNDSGSSRELNLVTGTIPVYRGNIYNI 60
 Db 1 MMSKYKRDLTVRQTVNVIAMKYDKLPVLDYVFNDSGSSRELNLVTGTIPVYRGNIYNI 60

QY 61 PICLWLLDTYPNPPICFVKPTSSMTIKTKHVDANGKIYLPYLDHDKHPRSELLELIQI 120
 Db 61 PICLWLLDTYPNPPICFVKPTSSMTIKTKHVDANGKIYLPYLDHDKHPRSELLELIQI 120

QY 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180
 Db 121 MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180

FT Modified-site 86 /note= "potential protein kinase C phosphorylation site"
 FT Modified-site 89 /note= "potential protein kinase C phosphorylation site"
 FT Domain 130..205 /note= "proline-rich domain"
 FT Modified-site 150 /note= "potential N-glycosylation site"
 FT Modified-site 210 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 215 /note= "potential protein kinase C phosphorylation site"
 FT Modified-site 225 /note= "potential protein kinase C phosphorylation site"
 FT Domain 231..302 /note= "coiled-coil domain"
 FT Modified-site 249 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 265 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 290 /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 297 /note= "potential N-glycosylation site"
 FT Modified-site 357 /note= "potential protein kinase C phosphorylation site"
 XX WO9718333-A1.
 XX 22-MAY-1997.
 XX 15-NOV-1996; 96WO-US018828.
 XX 16-NOV-1995; 95US-0006856P.
 XX 16-JAN-1996; 96US-00585758.
 XX 13-JUN-1996; 96US-00670274.
 XX (STRD) UNIV LELAND STANFORD JUNIOR.
 XX Li L, Cohen SN;
 XX WPI; 1997-289308/26.
 XX N-PSDB; AAT69575.
 XX Nucleic acid encoding tumour susceptibility protein TSG101 - useful to
 FT develop products for cancer detection, prophylaxis, therapy and
 FT predisposition determination.
 XX Example 1; Page 73-74; 89pp; English.
 XX This polypeptide sequence comprises a human tumour susceptibility
 CC protein, designated TSG101, that can function as a tumour suppressor in
 CC oncogenesis. Its amino acid sequence was deduced from a cDNA clone
 CC (AAT69575) isolated from a human placenta cDNA library. It shows 94%
 CC identity to human TSG101 (see AAW19110). Deletions in human TSG101 are
 CC associated with the occurrence of cancers, e.g. breast cancer. Methods
 CC are provided for identifying homologous or related proteins, for
 CC producing compositions that modulate the expression or function of the
 CC TSG101 protein, and for studying physiological pathways. Modulation of
 CC gene activity in vivo is used for prophylactic or therapeutic purposes,
 CC such as cancer therapy, identification of cell type based on expression,
 CC etc
 XX Sequence 380 AA;
 XX
 Query Match 95.2%; Score 1900.5; DB 2; Length 380;
 Best Local Similarity 94.5%; Pred. No. 3e-138;
 Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
 QY 1 MWSKXKVDLTIVROTVNVIAMVKDKPVLDSVFNDSGSSRELVLNLTGTIPVYRGNIYNI 60
 DB 1 MWSKXKVDLTIVROTVNVIAMVKDKPVLDSVFNDSGSSRELVLNLTGTIPVYRGNIYNI 60

QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGKVDANGKIYLPYLHDWKHPRSELLEIQI 120
 DB 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGKVDANGKIYLPYLHDWKHPRSELLEIQI 120
 QY 121 MIVIFGEBPPVSRPVSASYPPTATGPNTSYMPGMSGISAYPSGYPNPSGYPGCP 180
 DB 121 MIVVFGDEPPVFSRP- ISASYPPTATGPNTSYMPGMSGISAYPSGYPNPSGYPGCP 179
 QY 181 YPDAGYPATTSQYPSQPPVTVTPSRDGTISEDITRASLISAVSDKLRWRMKEEMDGA 240
 DB 180 YPDGGYPATTSQYPSQPPVTVTPSRDGTISEDITRASLISAVSDKLRWRMKEEMDGA 239
 QY 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMNQS 300
 DB 240 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMNQS 299
 QY 301 ENNDIDEVIPTAPLYKQILNLYAENAEEDTTFYLGELARRGVLDLDFELKHVRLLSRK 360
 DB 300 ENNDIDEVIPTAPLYKQILNLYAENAEEDTTFYLGELARRGVLDLDFELKHVRLLSRK 359
 QY 361 QFOLRALMOKARKTAGLSDIY 381
 DB 360 QFOLRALMOKARKTAGLSDIY 380
 RESULT 8
 AAW93424
 ID AAW93424 standard; protein; 390 AA.
 XX AC AAW93424;
 XX 11-JUN-1999 (first entry)
 XX Human NHTS protein.
 XX Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
 KW brain; breast; colon; heart; kidney; ovary; paraganglia; pancreas;
 KW prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
 KW biliary cirrhosis; Crohn's disease; rheumatoid arthritis.
 XX Homo sapiens.
 XX US5892016-A.
 XX 06-APR-1999.
 XX 23-JAN-1997; 97US-00786999.
 XX 23-JAN-1997; 97US-00786999.
 XX (INCY-) INCYTE PHARM.
 XX Brie SL, Goli SK;
 XX WPI; 1999-253932/21.
 XX N-PSDB; AAX23168.
 XX Novel human tumor suppressor - useful for the diagnosis or treatment of
 PT lymphoma, cancer, and autoimmune disease.
 XX Claim 1; Fig 1A-B; 25pp; English.
 XX This invention describes the isolation of a novel human tumour suppressor
 CC (NHTS). The products of the invention may be used for the diagnosis or
 CC treatment of conditions and diseases which are associated with expression
 CC of NHTS e.g. lymphoma and cancers of the brain, breast, colon, heart,
 CC kidney, ovary, paraganglia, pancreas, prostate, skin, stomach and thyroid
 CC and autoimmune disease e.g. asthma, biliary cirrhosis, Crohn's disease,
 CC diabetes, and rheumatoid arthritis
 XX Sequence 390 AA;
 XX Query Match 95.2%; Score 1900.5; DB 2; Length 390;

PN US6472508-BL.
XX 29-OCT-2002.
XX 18-DEC-1998; 98US-00216387.
XX 23-JAN-1997; 97US-00786999.
XX (INCY-) INCYTE GENOMICS INC.
XX La Brie S, Goli SK;
XX WPI; 2003-196850/19.
XX N-PSDB; ABQ77228.
XX New purified human tumor suppressor polypeptide, useful for diagnosing,
PT preventing and treating cancer and autoimmune diseases (e.g. asthma or
PT diabetes), and in drug screening assays.
XX
PS Claim 1; Fig 1A-B; 26pp; English.
XX
XX This invention describes a novel human tumor suppressor, NHTS, which has
CC cytostatic, antiasthmatic, hepatotropic, antiinflammatory, antidiabetic,
CC antirheumatic and antiarthritic activity. The NHTS polypeptide is useful
CC in diagnosing, preventing and treating cancer and autoimmune diseases
CC (e.g. asthma, biliary cirrhosis, Crohn's disease, diabetes or rheumatoid
CC arthritis), in drug screening assays, in vaccines and for gene therapy.
CC The nucleic acid encoding the above polypeptide may be used to detect and
CC quantitative gene expression in biopsied tissues in which expression of the
CC polypeptide may be correlated with disease, and thus, may be used to
CC distinguish between absence, presence, and excess expression of the
CC polypeptide, and to monitor the regulation of the polypeptide levels
CC during therapeutic intervention. The polynucleotide may also be used to
CC generate hybridisation probes for mapping genomic sequences, and in
CC various molecular biology techniques. Antibodies generated from the
CC products of the invention may be used as a targeting or delivery
CC mechanism for bringing an agonist or other pharmaceutical agent to cells
CC of patients having the above-mentioned diseases, and in diagnosing
CC conditions or diseases characterised by the expression of the
CC polypeptide, or in assays to monitor patients being treated with the
CC above polypeptide, agonists or antagonists. This sequence represents the
CC NHTS protein, isolated from Incyte clone 609476 derived from COLNNOT01, a
CC colon CDNA library
XX
SQ Sequence 390 AA;
Query Match 95.2%; Score 1900.5; DB 6; Length 390;
Best Local Similarity 94.5%; Pred. No. 3.1e-138;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MMSYKYRDLTVROTAVNIAMKDKPVLDYVFNDSGSSRELNLVLTGTIPVYRGNIYNI 60
DB 11 MVSXYKYRDLTVRETAVNITLYKDKPVLDYVFNDSGSSRELNLVLTGTIPVYRGNTYNI 70
QY 61 PICLLWLLDTPYNPPICFVKPTSSMTIKTGKHDVANGKIYLYPLHDKWHPKRSLELLIQI 120
DB 71 PICLLWLLDTPYNPPICFVKPTSSMTIKTGKHDVANGKIYLYPLHDKWHPKRSLELLIQI 130
QY 121 MIVIFGEBPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPPNPSGPGCP 180
DB 131 MIVVFGDEPPVFSRP-ISASYPPTYQATGPPNTSYMPGMPGSGISAYPSGYPPNPSGPGCP 189
QY 181 YPPAGPYPATSSQYPSQPPVTTVGPSSDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
DB 190 YPPGGPYPATSSQYPSQPPVTTVGPSSDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 249
QY 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEMENQS 300
DB 250 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEMENQS 309
QY 301 ENNDIDEVIIPTAPLYKQILNLYAEENAEITDFYLGALRRGVLDLDFLKHVLLSRK 360
DB 310 ENNDIDEVIIPTAPLYKQILNLYAEENAEITDFYLGALRRGVLDLDFLKHVLLSRK 369

QY 361 QFQIRALMOKARKTAGLSDL 381
DB 370 QFQIRALMOKARKTAGLSDL 390
RESULT 11
ADC35176
ID ADC35176 standard; protein; 390 AA.
XX ADC35176;
XX 18-DEC-2003 (first entry)
XX Novel human tumor suppressor, NHTS.
XX novel human tumor suppressor; NHTS; human; tumor; cancer;
KW connective tissue disorder; rheumatoid arthritis.
XX Homo sapiens.
XX OS
XX US2003099988-A1.
XX 29-MAY-2003.
XX 13-SEP-2002; 2002US-00243815.
XX 23-JAN-1997; 97US-00786999.
XX 18-DEC-1998; 98US-00216387.
XX (INCY-) INCYTE GENOMICS INC.
XX PA
XX Brie SL, Goli SK;
XX WPI; 2003-687760/65.
XX N-PSDB; ADC35177, ADC35179.
XX Novel human tumor suppressor (NHTS) useful for preparing a composition
PT for diagnosing, treating or preventing a disease or condition associated
PT with expression of NHTS e.g. cancer.
XX
PS Claim 1; SEQ ID NO 1; 29pp; English.
XX
XX The invention relates to an isolated polypeptide. The polypeptide is
CC useful for preparing a composition for diagnosing or treating a disease
CC or condition associated with decreased expression or overexpression of
CC novel human tumor suppressor (NHTS) e.g. cancer or connective tissue
CC disorders, such as rheumatoid arthritis. The present sequence represents
CC the amino acid sequence of novel human tumor suppressor, NHTS.
XX
SQ Sequence 390 AA;
Query Match 95.2%; Score 1900.5; DB 7; Length 390;
Best Local Similarity 94.5%; Pred. No. 3.1e-138;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MMSYKYRDLTVROTAVNIAMKDKPVLDYVFNDSGSSRELNLVLTGTIPVYRGNIYNI 60
DB 11 MVSXYKYRDLTVRETAVNITLYKDKPVLDYVFNDSGSSRELNLVLTGTIPVYRGNTYNI 70
QY 61 PICLLWLLDTPYNPPICFVKPTSSMTIKTGKHDVANGKIYLYPLHDKWHPKRSLELLIQI 120
DB 71 PICLLWLLDTPYNPPICFVKPTSSMTIKTGKHDVANGKIYLYPLHDKWHPKRSLELLIQI 130
QY 121 MIVIFGEBPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPPNPSGPGCP 180
DB 131 MIVVFGDEPPVFSRP-ISASYPPTYQATGPPNTSYMPGMPGSGISAYPSGYPPNPSGPGCP 189
QY 181 YPPAGPYPATSSQYPSQPPVTTVGPSSDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
DB 190 YPPGGPYPATSSQYPSQPPVTTVGPSSDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 249
QY 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQEAEDVKNIELLKKKDEELSSALEMENQS 300

```
Db 250 QAEINLAKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNIELLKKDEELSSALEKMNQS 309
QY 301 ENNDIDEVIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 310 ENNDIDEVIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 369
QY 361 QFOLRALMOKARKTAGLSLDLY 381
Db 370 QFOLRALMOKARKTAGLSLDLY 390

RESULT 12
AAE34885
ID AAE34885 standard; protein; 390 AA.
XX
AC AAE34885;
XX
DT 28-MAY-2003 (first entry)
XX
DE Human Tsg101 protein.
XX
KW Human; twisted gastrulation; Tsg101; human immunodeficiency virus; HIV;
KW gene therapy; peptide therapy.
XX
OS Homo sapiens.
XX
EN WO200294314-A1.
XX
PD 28-NOV-2002.
XX
PF 21-MAY-2002; 2002WO-US015965.
XX
PR 21-MAY-2001; 2001US-0292761P.
XX
PA (UUNY ) UNIV NEW YORK STATE RES FOUND.
XX
PI Cohen SN, Carter C, Goff A, Ehrlich L;
XX WPI; 2003-148440/14.
XX
PT Identifying twisted gastrulation 101 peptide, for treating human
PT immunodeficiency virus (HIV) infection, comprises comparing the level of
PT HIV viral particles in a mammalian cell culture to that in a control
PT culture.
XX
PS Claim 9; Col 45; 35pp; English.
XX
CC The invention relates to a method of identifying a mammalian twisted
CC gastrulation (Tsg) 101 peptide. The method involves measuring the level
CC of human immunodeficiency virus (HIV) viral particles released in a
CC culture of mammalian cells having an expression construct comprising a
CC portion of the coding sequence of a mammalian Tsg101 gene and comparing
CC the level of HIV viral particles to that in a culture of control
CC mammalian cells. The method is useful in identifying a peptide that is
CC effective in reducing HIV particle production or which may be used in
CC treating a patient infected with HIV or other retrovirus. The invention
CC is useful in gene therapy and peptide therapy. The present sequence is
CC human Tsg101 protein
XX
SQ Sequence 390 AA;

Query Match 95.0%; Score 1895.5; DB 6; Length 390;
Best Local Similarity 94.2%; Pred. No. 7.5e-138;
Matches 359; Conservative 11; Mismatches 10; Indels 1; Gaps 1;

QY 1 MMSKYKRDLTVRQTVNVNVIAMYKOLKPVLDGVFNDGSSRELNVLTGTIPYVRGNLYNI 60
Db 11 MVSXYKYRDLTVRETVNVITLYKDLKPVLDGVFNDGSSRELMLTGTIPYVRGNLYNI 70
QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDKHPRSELLELTQI 120
Db 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDKHPRSELLELTQI 130
```

```
QY 121 MIVIFGEPPVFSRPTVSASYPPTATGPNTSYMPGMPBSGISAYPSGYPPNPSGYPGCP 180
Db 131 MIVVFGDEPVPFSRP-ISASYPPTATGPNTSYMPGMPBSGISAYPSGYPPNPSGYPGCP 189
QY 181 YPPAGYPPTATSSQYPSQPPVTITVPSRDGTISEDITRASLISAVSDKLRWRMKEEMDGA 240
Db 190 YPPGGYPPTATSSQYPSQPPVTITVPSRDGTISEDITRASLISAVSDKLRWRMKEEMDRA 249
QY 241 QAEINLAKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNIELLKKDEELSSALEKMNQS 300
Db 250 QAEINLAKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNIELLKKDEELSSALEKMNQS 309
QY 301 ENNDIDEVIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 310 ENNDIDEVIPTAPLYKQILNLNLYAEENAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 369
QY 361 QFOLRALMOKARKTAGLSLDLY 381
Db 370 QFOLRALMOKARKTAGLSLDLY 390

RESULT 13
ABP41729
ID ABP41729 standard; protein; 307 AA.
XX
AC ABP41729;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HSPSE88, SEQ ID NO:2861.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 11p15.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US018569.
XX
PR 07-JUN-2000; 2000US-0209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX WPI; 2002-147878/19.
XX
DR N-PSDB; ABQ54806.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT cancer), immune disorders, cardiovascular disorders and neurological
PT diseases.
XX
PS Claim 11; SEQ ID NO 2861; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
```

CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), and toxic
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and oophoritis and
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ

Query Match 67.0%; Score 1337.5; DB 5; Length 307;
Best Local Similarity 95.2%; Pred. No. 6.6e-95;
Matches 259; Conservative 6; Mismatches 6; Indels 1; Gaps 1;
QY 110 PRSELLELIQIMIVFGEPPVFSRPTVSASYPYATATGPPNTSYMPGSGISAYPSGY 169
DB 37 PQSDLLGLIQIMIVFGEPPVFSRPTVSASYPYATATGPPNTSYMPGSGISAYPSGY 95
QY 170 PPNPSGPGCPYPAGYPATTSSQYPSQPPVTWGPSSRDTISSEDITIRASLISAVSDKL 229
DB 96 PPNPSGPGCPYPAGYPATTSSQYPSQPPVTWGPSSRDTISSEDITIRASLISAVSDKL 155
QY 230 RWRKKEEMDGAQELNALKRTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKDEEL 289
DB 156 RWRKKEEMDGAQELNALKRTEEDLKKGHOKLEEMVTRLDQEAEDVKNIELKKDEEL 215
QY 290 SSALEKMEQSENNIDIEVIITAPLYKQILNLVAENAIEDTIFYLGEALLRRGVLDLV 349
DB 216 SSALEKMEQSENNIDIEVIITAPLYKQILNLVAENAIEDTIFYLGEALLRRGVLDLV 275
QY 350 FLKHVRLLSRKQFOLRALMQKARTAGLSDIY 381
DB 276 FLKHVRLLSRKQFOLRALMQKARTAGLSDIY 307

RESULT 14
AAE09328
ID AAE09328 standard; protein; 237 AA.
XX
AC AAE09328;
XX
DT 19-NOV-2001 (first entry)
XX Human intracellular regulatory molecule, tsig101.
XX
KW Human; intracellular regulator; cell division; proliferation; therapy;
KW cancer; infection; wound; developmental abnormality; metabolic problem;
KW cytostatic; antibacterial; vulnerary; transcription factor; tsig101.
XX
OS Homo sapiens.
XX
FN US6274312-B1.
XX
PD 14-AUG-2001.
XX
PF 10-DEC-1997; 97US-00999774.
XX

PR 11-DEC-1996; 96US-0032818P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Gish KC, Seghezzi W, Shanahan F, Lees EM, Mcclanahan TK;
XX
DR WPI; 2001-535086/59.
XX N-PSDB; AAD16104.
DR
XX
PT New genes encoding intracellular regulatory molecules, useful for
PT regulating cell division and proliferation (e.g. tumor cells),
PT particularly for treating cancer, infections, wounds, or developmental or
PT metabolic abnormalities.
XX
PS Example 2; Col 29-32; 62pp; English.
XX
CC The present sequence is a human intracellular regulatory molecule,
CC tsig101. The tsig101 is a transcription factor. The polynucleotides
CC encoding intracellular regulatory molecules are useful for regulating
CC cell division and proliferation of various cell types, including tumour
CC cells. Specifically, they are also useful for treating cancer,
CC infections, wounds, developmental abnormalities or metabolic problems
XX
SQ Sequence 237 AA;

Query Match 51.2%; Score 1022.5; DB 4; Length 237;
Best Local Similarity 88.3%; Pred. No. 9.9e-71;
Matches 182; Conservative 12; Mismatches 11; Indels 1; Gaps 1;
QY 1 MMSKYYKRDLTQVQNVNVIAMWKDLKPVLDYVFDGSSRELNLVTCTIPVRKGNLYNI 60
DB 11 MVSXKYYKRDLTQVQNVNVIAMWKDLKPVLDYVFDGSSRELNLVTCTIPVRKGNLYNI 70
QY 61 PICLWLDTPYNNPPICFVKPTSSMTITKTKHVDANGKIYLPYLHDKHPRSELLELIQI 120
DB 71 PICLWLDTPYNNPPICFVKPTSSMTITKTKHVDANGKIYLPYLHDKHPRSELLELIQI 130
QY 121 MIVFGEPPVFSRPTVSASYPYATATGPPNTSYMPGSGISAYPSGYPNPSPGPGCP 180
DB 131 MIVFGEPPVFSRPTVSASYPYATATGPPNTSYMPGSGISAYPSGYPNPSPGPGCP 189
QY 181 YPPAGYPATTSSQYPSQPPVTWGP 206
DB 190 YPPAGYPATTSSQYPSQPPVTWGP 215

RESULT 15
ABB64607
ID ABB64607 standard; protein; 331 AA.
XX
AC ABB64607;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 20613.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
FN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;

PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	28-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147935P.
PR	09-AUG-1999;	99US-0148171P.
PR	10-AUG-1999;	99US-0148319P.
PR	11-AUG-1999;	99US-0148341P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	22-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161933P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 22.0%; Score 440; DB 3; Length 398;		
Best Local Similarity 31.6%; Pred. No. 1.9e-25;		
Matches 117; Conservative 72; Mismatches 147; Indels 34; Gaps 12;		
Qy	12 VRQ-TVNVIAMYKDLKPVLDYVNDGSSRELNLNLTGTIPVRYEGNTYNIPICLWLDDTY	70
Db	37 IRQHLLNLSIYSPSLEPKTASFMHNDGRSVNLLQADGTIPMPFHGVTYNIPIVILLESY	96
Qy	71 PYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEP	129
Db	97 PRHPFCVYNPTADMIKRPFAHVTPSGLYSLPLQNVVYPSSNLVDLSLSAFAFADP	156
Qy	130 PVFSRPTVSASYPPTATGPNT--SYMPGMPGSIASYPSPNPSPGCPYPPAGPY	187
Db	157 PLYSR--RRPQP--PPSPPTVYDSSLRPPSADQSLRPFPPSPYG-----GGVS	203
Qy	188 PATSSQYPSQPPVTVTPSPEDGTISE--DTIRASLISAVSDKLRWKMKEMDGAQELN	245
Db	204 RVQVQHVOHQOQSDAAEVFKRNAINKWNEMVHSDLS--MRRAREAEELLSLQA---	258
Qy	246 ALKTEEDLKKGHQKLEBMTRLDQEAVIDKNEILLKKDKDELSALEMENQSEN---	302
Db	259 GLKREDELNIG---LKEMV---EKKETLEQQQLIISMTNTDILDSWVRENOGKTNLVD	311
Qy	303 NDIEVIPTAPLYKQILNLYAENAIETDTIFYLGEALRGVIDLDVFLKHVRLLSKQF	362
Db	312 LDVDNAFECGDTLSKQMLECTALDIAIEDAIYSLDKSFQDGVVFPDQYLRNVRLLSRQF	371
Qy	363 QLRLMOKAR	372
Db	372 PHRATGSKVR	381
RESULT 17		
AAG06369		
ID AAG06369 standard; protein; 412 AA.		
XX		

AC AAG06369; AC 18-JUN-1999; 99US-0139463P.
XX DT 17-OCT-2000 (first entry) PR 18-JUN-1999; 99US-0139750P.
XX XX 99US-0139763P. PR 18-JUN-1999; 99US-0139817P.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 3118. PR 21-JUN-1999; 99US-0139899P.
XX XX 99US-0140353P. PR 22-JUN-1999; 99US-0140354P.
KW Protein identification; signal transduction pathway; metabolic pathway; PR 23-JUN-1999; 99US-0140354P.
KW hybridisation assay; genetic mapping; gene expression control; promoter; PR 23-JUN-1999; 99US-0140695P.
KW termination sequence. PR 24-JUN-1999; 99US-0140823P.
XX XX 99US-0140991P. PR 28-JUN-1999; 99US-0140991P.
OS Arabidopsis thaliana. PR 29-JUN-1999; 99US-0141287P.
XX XX 99US-0141842P. PR 30-JUN-1999; 99US-0141842P.
XX PN EP1033405-A2. PR 01-JUL-1999; 99US-0142154P.
XX XX 06-SEP-2000. PR 01-JUL-1999; 99US-0142055P.
XX PD 99US-0142154P. PR 02-JUL-1999; 99US-0142390P.
XX PF 25-FEB-2000; 2000EP-00301439. PR 06-JUL-1999; 99US-0142803P.
XX XX 99US-0142920P. PR 09-JUL-1999; 99US-0142920P.
XX XX 99US-0142977P. PR 12-JUL-1999; 99US-0142977P.
PR 25-FEB-1999; 99US-0121825P. PR 13-JUL-1999; 99US-0143542P.
PR 05-MAR-1999; 99US-0123180P. PR 14-JUL-1999; 99US-0143624P.
PR 09-MAR-1999; 99US-0123548P. PR 15-JUL-1999; 99US-0144005P.
PR 23-MAR-1999; 99US-0125788P. PR 16-JUL-1999; 99US-0144085P.
PR 25-MAR-1999; 99US-0126264P. PR 16-JUL-1999; 99US-0144085P.
PR 29-MAR-1999; 99US-0126785P. PR 19-JUL-1999; 99US-0144325P.
PR 01-APR-1999; 99US-0127462P. PR 19-JUL-1999; 99US-0144331P.
PR 06-APR-1999; 99US-0128234P. PR 19-JUL-1999; 99US-0144332P.
PR 08-APR-1999; 99US-0128714P. PR 19-JUL-1999; 99US-0144333P.
PR 16-APR-1999; 99US-0129845P. PR 19-JUL-1999; 99US-0144334P.
PR 19-APR-1999; 99US-0130077P. PR 19-JUL-1999; 99US-0144335P.
PR 21-APR-1999; 99US-0130449P. PR 20-JUL-1999; 99US-0144352P.
PR 23-APR-1999; 99US-0130510P. PR 20-JUL-1999; 99US-0144884P.
PR 23-APR-1999; 99US-0130891P. PR 21-JUL-1999; 99US-0145086P.
PR 28-APR-1999; 99US-0131449P. PR 21-JUL-1999; 99US-0145088P.
PR 30-APR-1999; 99US-0132048P. PR 22-JUL-1999; 99US-0145088P.
PR 30-APR-1999; 99US-0132407P. PR 22-JUL-1999; 99US-0145089P.
PR 04-MAY-1999; 99US-0132484P. PR 22-JUL-1999; 99US-0145089P.
PR 05-MAY-1999; 99US-0132485P. PR 23-JUL-1999; 99US-0145192P.
PR 06-MAY-1999; 99US-0132486P. PR 23-JUL-1999; 99US-0145145P.
PR 06-MAY-1999; 99US-0132487P. PR 23-JUL-1999; 99US-0145218P.
PR 07-MAY-1999; 99US-0132488P. PR 23-JUL-1999; 99US-0145224P.
PR 07-MAY-1999; 99US-0132863P. PR 26-JUL-1999; 99US-0145276P.
PR 11-MAY-1999; 99US-0134256P. PR 27-JUL-1999; 99US-0145913P.
PR 14-MAY-1999; 99US-0134218P. PR 27-JUL-1999; 99US-0145918P.
PR 14-MAY-1999; 99US-0134219P. PR 27-JUL-1999; 99US-0145919P.
PR 14-MAY-1999; 99US-0134221P. PR 28-JUL-1999; 99US-0145951P.
PR 14-MAY-1999; 99US-0134370P. PR 02-AUG-1999; 99US-0146386P.
PR 18-MAY-1999; 99US-0134768P. PR 02-AUG-1999; 99US-0146388P.
PR 19-MAY-1999; 99US-0134941P. PR 02-AUG-1999; 99US-0146389P.
PR 20-MAY-1999; 99US-0135124P. PR 02-AUG-1999; 99US-0146389P.
PR 21-MAY-1999; 99US-0135153P. PR 03-AUG-1999; 99US-0147038P.
PR 24-MAY-1999; 99US-0135629P. PR 03-AUG-1999; 99US-0147038P.
PR 25-MAY-1999; 99US-0136021P. PR 04-AUG-1999; 99US-0147204P.
PR 27-MAY-1999; 99US-0136392P. PR 04-AUG-1999; 99US-0147302P.
PR 28-MAY-1999; 99US-0136782P. PR 05-AUG-1999; 99US-0147192P.
PR 01-JUN-1999; 99US-0137222P. PR 05-AUG-1999; 99US-0147192P.
PR 03-JUN-1999; 99US-0137528P. PR 05-AUG-1999; 99US-0147260P.
PR 04-JUN-1999; 99US-0137502P. PR 06-AUG-1999; 99US-0147303P.
PR 07-JUN-1999; 99US-0137724P. PR 06-AUG-1999; 99US-0147416P.
PR 08-JUN-1999; 99US-0138094P. PR 09-AUG-1999; 99US-0147935P.
PR 10-JUN-1999; 99US-0138540P. PR 09-AUG-1999; 99US-0147935P.
PR 10-JUN-1999; 99US-0138847P. PR 10-AUG-1999; 99US-0148171P.
PR 14-JUN-1999; 99US-0139119P. PR 11-AUG-1999; 99US-0148319P.
PR 16-JUN-1999; 99US-0139452P. PR 12-AUG-1999; 99US-0148341P.
PR 16-JUN-1999; 99US-0139453P. PR 13-AUG-1999; 99US-0148565P.
PR 17-JUN-1999; 99US-0139492P. PR 16-AUG-1999; 99US-0149368P.
PR 18-JUN-1999; 99US-0139454P. PR 17-AUG-1999; 99US-0149175P.
PR 18-JUN-1999; 99US-0139455P. PR 18-AUG-1999; 99US-0149426P.
PR 18-JUN-1999; 99US-0139456P. PR 20-AUG-1999; 99US-0149722P.
PR 18-JUN-1999; 99US-0139457P. PR 20-AUG-1999; 99US-0149723P.
PR 18-JUN-1999; 99US-0139458P. PR 20-AUG-1999; 99US-0149929P.
PR 18-JUN-1999; 99US-0139459P. PR 23-AUG-1999; 99US-0149902P.
PR 18-JUN-1999; 99US-0139460P. PR 23-AUG-1999; 99US-0149930P.
PR 18-JUN-1999; 99US-0139461P. PR 25-AUG-1999; 99US-0150566P.
PR 18-JUN-1999; 99US-0139462P.

PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.0%; Score 440; DB 3; Length 412;
Best Local Similarity 31.6%; Pred. No. 2e-25;
Matches 117; Conservative 72; Mismatches 147; Indels 34; Gaps 12;

QY 12 VRQ-TVNVIAMKLPVLSYFVNDGSSRELVLNLTGTIPVYRGNIYNIPICLWLDITY 70
DB 51 IROHLLNLISYPSELEPKTASFMNDGSRVNLQADGTIPMPFHGVTNIPVILWLESY 110
QY 71 PYNPPICVKTSSMTIK-TCKHYVDANGKIYLPYLHWKHPRSSELLELIQIMVIFGEEP 129
DB 111 PRHPFCVNVNTADMIKRPAAHTVPSGLVSLPLQNVWVYSSNLVDLSAFAFADP 170
QY 130 PVFGRPTVASYPYATGPNWT--SYMPGMPGISAVPSGYPNPSPGCPYPPAGPY 187
DB 171 PLYSR--RRPQP--PPSPPTVYDSSLRPPSDQSLRFPFPSPYG-----GGVS 217
QY 188 PATTSSQYPSOPPVTTVGPRDGTISE--DTIRASLISAVSDKLRWRMKEEMDGAQALN 245
DB 218 RVQVQVHHQOQSDAAAEVFKRNAINKMWVHSDLV5--MRRAREABAEELLSLOA--- 272

QY 246 ALKRTEDLKKGHKLEEMVTRLDQVAEVVDKNIELKKDKDEELSSALLEKMEMNOSEN--- 302
DB 273 GLARREDELNIG---LKEMV---EKETLEQQLIISMNTDILDSWVRENQGTKNLVD 325
QY 303 NDIDEVITPTAPLYKQILNLYAEENAIEDTIFYLGEALRGVIDLDVFLKHLVLLSKQF 362
DB 326 LDVDNAFECGDTLSKQMLECTALDLAIEDAIVSLDKSFQDGVVFFDQYLRNVLSSREQF 385
QY 363 QLRALMQKAR 372
DB 386 FHRATGSKVR 395
RESULT 18
ABP43551
ID ABP43551 standard; protein; 379 AA.
XX
AC ABP43551;
XX
XX 26-FEB-2003 (first entry)
XX
DE L-lactate dehydrogenase M-chain protein.
XX
XX Neuroprotective; immunomodulator; cancer; chromosome 11; cytostatic;
KW anti-inflammatory; gene therapy; nutritional supplement; wound; burn;
KW ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnerary.
XX
OS Homo sapiens.
XX
XX WC200231111-A2.
XX
PD 18-APR-2002.
XX
XX 11-OCT-2001; 2001WO-US027760.
XX
XX 12-OCT-2000; 2000US-00687527.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX WPI: 2002-426278/45.
DR N-PSDB; ABQ60795.
XX
XX New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.
XX
XX Claim 20; SEQ ID # 454; 357pp + Sequence Listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnerary, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABP43544-
CC ABP43989 represent polypeptides encoded by polynucleotides of the
CC invention. NOTE: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 379 AA;
Query Match 20.8%; Score 415; DB 5; Length 379;
Best Local Similarity 54.9%; Pred. No. 1.5e-23;

PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	09-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match		
20.4%; Score 406.5; DB 3; Length 340;		
Best Local Similarity 31.5%; Pred. No. 6e-23;		
Matches 109; Conservative 65; Mismatches 139; Indels 33; Gaps 11;		
QY	35	NDGSSRELNLGTIPVRVREGNIYINPICLWLIDTYPPNPPICFVAKTSSMTIK-TGKHV 93
Db	3	NDGRSVNLLQADGTIPMPFHGVTYINPIVILWLESYPRHPPCVYVNPNTADMIIRPHAHV 62
QY	94	DANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEPPVFRSPTVSASYPPTATGPENT- 152
Db	63	TPSGLVSLPYLQNWVTPSSNLVDLSNAFARDEPLYSR---RRQPP--PPSPPTVY 117
QY	153	-SYMPGMGSGISAYSPGYPNPGYPCPYPPAGYPATTSQYPSQYPPVTTVGPSRDGT 211
Db	118	DSLSRPPSADQSLPRFPSPYGVG-----GGVSRVQVQHVHHQQQSDDAAFVKRNA 169
QY	212	ISE--DTIRASILSAVSDKLRWKEBMDGAQELNALKETEDLKKGHQKLEBMTFLD 269
Db	170	INKAVENVHSDLVN--MRRAREAEALLSLQA---GLKRRDELNTG---LKEMV---- 217
QY	270	QEAEDVKNTELKKKDEELSSALEKMNQSEN---NDIDEVIPTAPLYKQILNLYAE 326
Db	218	BEKETLEQQLQIISMNTDILDSWRENQKTKNLVDLDVDFNAFECGDTLSKOMLECTALD 277
QY	327	NAIEDTIFYLGEALRGVVIDLDFLKHVLLSRKQFOLRALMQXAR 372
Db	278	LATIEDALYSLDKSGFDQGVWPFQVLRNVRLLSREQPFPHRATGSKVR 323
RESULT 20		
AAG01689		
ID	AAG01689 standard; protein; 90 AA.	
AC	AAG01689;	
XX	06-OCT-2000 (first entry)	
DE	Human secreted protein, SEQ ID NO: 5770.	
XX	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;	
KW	gene therapy; chromosome mapping.	
OS	Homo sapiens.	
XX	EP1033401-A2.	
FN	06-SEP-2000.	
PD	21-FEB-2000; 2000EP-00200610.	
XX	26-FEB-1999; 99US-0122487P.	
PF	(GEST) GENSET.	
XX	Dumas Milne Edwards J, Duclert A, Giordano J;	
PI	WPI; 2000-500381/45.	
XX	N-PSDB; AAC01695.	
DR	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
XX	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures.	
XX	Claim 13; SEQ ID NO 5770; 71pp + Sequence Listing; English.	
PS	The present sequence is a polypeptide encoded by one of a large number of	
XX	5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were	
CC	prepared from total human RNAs or polyA+ RNAs derived from 30 different	
CC	tissues. EST sequences usually correspond mainly to the 3' untranslated	
CC	region (UTR) of the mRNA because they are often obtained from oligo-dT	
CC	primed cDNA libraries. Such ESTs are not well suited for isolating cDNA	
CC	sequences derived from the 5' ends of mRNAs and even in those cases where	
CC	longer cDNA sequences have been obtained, the full 5' UTR is rarely	
CC	included. 5' ESTs are derived from mRNAs with intact 5' ends and can	

CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors

XX Sequence 90 AA;

Query Match 20.1%; Score 402; DB 3; Length 90;
Best Local Similarity 91.2%; Pred. No. 2.2e-23;
Matches 73; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 MMSKYRDLTYRQTVNVIAMVKDLKPVLDYVNDGSSRELNLVTGTTPVYRGNIYNI 60
Db 11 MMSKYRDLTYRQTVNVIAMVKDLKPVLDYVNDGSSRELNLVTGTTPVYRGNIYNI 70
QY 61 PICLWLLDTPYNPPICFVK 80
Db 71 PICLWLLDTPYNPPICFVK 90

RESULT 21

AAG13898
ID AAG13898 standard; protein; 322 AA.

XX AAG13898;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 13559.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 23-MAR-1999; 99US-0123548P.

PR 25-MAR-1999; 99US-0125788P.

PR 29-MAR-1999; 99US-0126264P.

PR 01-APR-1999; 99US-0126785P.

PR 06-APR-1999; 99US-0127462P.

PR 08-APR-1999; 99US-0128234P.

PR 16-APR-1999; 99US-0128714P.

PR 19-APR-1999; 99US-0129845P.

PR 21-APR-1999; 99US-0130077P.

PR 23-APR-1999; 99US-0130449P.

PR 28-APR-1999; 99US-0130510P.

PR 30-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0134449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 05-MAY-1999; 99US-0132484P.

PR 06-MAY-1999; 99US-0132485P.

PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.

PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241788P.
PR 20-OCT-2000; 2000US-0241789P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465566/50.
XX N-PSDE; AAS40986.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,
XX treating neural, immune system, muscular, reproductive, pulmonary,
XX cardiovascular, renal, proliferative disorders and cancerous diseases.
XX
XX Claim 11; SEQ ID NO 1112; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders including hyperproliferative disorders (e.g. cancer),

CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
 CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
 CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
 CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
 CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
 CC infectious disorders (e.g. Influenza). The polynucleotides of the
 CC invention can also be used in gene therapy. AAU22915-AAU23814 represent
 CC the novel human enzyme polypeptides of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 452 AA;
 SQ

Query Match 15.9%; Score 318; DB 4; Length 452;
 Best Local Similarity 45.1%; Pred. No. 6e-16;
 Matches 60; Conservative 21; Mismatches 30; Indels 22; Gaps 1;

QY 1 MMSKKYRDITVROVNVVIAWYKDLKPVLDYVDFGSSRELNVLTGTPVYRGNIYNI 60
 DB 14 LKGYKFRDLTVEELRNVNVPFPHFKYSMDTY-----GNTYNI 51
 QY 61 PICLWLDTPYNNPICFVKPTSSMTIKGKHVDANGKIYLPYLDHDKHPRSELLELIQI 120
 DB 52 PIRFWLDHPAPPICFLKPTANMGLVKGHVDAGRIYLPYLDHDKHPRSELLELIQI 111
 QY 121 MIVIFGEPPPVES 133
 DB 112 MIAKQEELEPMYS 124

RESULT 25
 AAG13899
 ID AAG13899 standard; protein; 288 AA.
 XX
 AC AAG13899;
 XX
 DT 17-OCT-2000 (first entry)
 DE
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 13560.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 29-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0130891P.
 PR 30-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 06-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132488P.
 PR 11-MAY-1999; 99US-0132489P.
 PR 14-MAY-1999; 99US-0132490P.
 PR 14-MAY-1999; 99US-0132491P.
 PR 14-MAY-1999; 99US-0132492P.
 PR 14-MAY-1999; 99US-0132493P.
 PR 14-MAY-1999; 99US-0132494P.
 PR 14-MAY-1999; 99US-0132495P.
 PR 14-MAY-1999; 99US-0132496P.
 PR 14-MAY-1999; 99US-0132497P.
 PR 14-MAY-1999; 99US-0132498P.
 PR 14-MAY-1999; 99US-0132499P.
 PR 14-MAY-1999; 99US-0132500P.
 PR 14-MAY-1999; 99US-0132501P.
 PR 14-MAY-1999; 99US-0132502P.
 PR 14-MAY-1999; 99US-0132503P.
 PR 14-MAY-1999; 99US-0132504P.
 PR 14-MAY-1999; 99US-0132505P.
 PR 14-MAY-1999; 99US-0132506P.
 PR 14-MAY-1999; 99US-0132507P.
 PR 14-MAY-1999; 99US-0132508P.
 PR 14-MAY-1999; 99US-0132509P.
 PR 14-MAY-1999; 99US-0132510P.
 PR 14-MAY-1999; 99US-0132511P.
 PR 14-MAY-1999; 99US-0132512P.
 PR 14-MAY-1999; 99US-0132513P.
 PR 14-MAY-1999; 99US-0132514P.
 PR 14-MAY-1999; 99US-0132515P.
 PR 14-MAY-1999; 99US-0132516P.
 PR 14-MAY-1999; 99US-0132517P.
 PR 14-MAY-1999; 99US-0132518P.
 PR 14-MAY-1999; 99US-0132519P.
 PR 14-MAY-1999; 99US-0132520P.
 PR 14-MAY-1999; 99US-0132521P.
 PR 14-MAY-1999; 99US-0132522P.
 PR 14-MAY-1999; 99US-0132523P.
 PR 14-MAY-1999; 99US-0132524P.
 PR 14-MAY-1999; 99US-0132525P.
 PR 14-MAY-1999; 99US-0132526P.
 PR 14-MAY-1999; 99US-0132527P.
 PR 14-MAY-1999; 99US-0132528P.
 PR 14-MAY-1999; 99US-0132529P.
 PR 14-MAY-1999; 99US-0132530P.
 PR 14-MAY-1999; 99US-0132531P.
 PR 14-MAY-1999; 99US-0132532P.
 PR 14-MAY-1999; 99US-0132533P.
 PR 14-MAY-1999; 99US-0132534P.
 PR 14-MAY-1999; 99US-0132535P.
 PR 14-MAY-1999; 99US-0132536P.
 PR 14-MAY-1999; 99US-0132537P.
 PR 14-MAY-1999; 99US-0132538P.
 PR 14-MAY-1999; 99US-0132539P.
 PR 14-MAY-1999; 99US-0132540P.
 PR 14-MAY-1999; 99US-0132541P.
 PR 14-MAY-1999; 99US-0132542P.
 PR 14-MAY-1999; 99US-0132543P.
 PR 14-MAY-1999; 99US-0132544P.
 PR 14-MAY-1999; 99US-0132545P.
 PR 14-MAY-1999; 99US-0132546P.
 PR 14-MAY-1999; 99US-0132547P.
 PR 14-MAY-1999; 99US-0132548P.
 PR 14-MAY-1999; 99US-0132549P.
 PR 14-MAY-1999; 99US-0132550P.
 PR 14-MAY-1999; 99US-0132551P.
 PR 14-MAY-1999; 99US-0132552P.
 PR 14-MAY-1999; 99US-0132553P.
 PR 14-MAY-1999; 99US-0132554P.
 PR 14-MAY-1999; 99US-0132555P.
 PR 14-MAY-1999; 99US-0132556P.
 PR 14-MAY-1999; 99US-0132557P.
 PR 14-MAY-1999; 99US-0132558P.
 PR 14-MAY-1999; 99US-0132559P.
 PR 14-MAY-1999; 99US-0132560P.
 PR 14-MAY-1999; 99US-0132561P.
 PR 14-MAY-1999; 99US-0132562P.
 PR 14-MAY-1999; 99US-0132563P.
 PR 14-MAY-1999; 99US-0132564P.
 PR 14-MAY-1999; 99US-0132565P.
 PR 14-MAY-1999; 99US-0132566P.
 PR 14-MAY-1999; 99US-0132567P.
 PR 14-MAY-1999; 99US-0132568P.
 PR 14-MAY-1999; 99US-0132569P.
 PR 14-MAY-1999; 99US-0132570P.
 PR 14-MAY-1999; 99US-0132571P.
 PR 14-MAY-1999; 99US-0132572P.
 PR 14-MAY-1999; 99US-0132573P.
 PR 14-MAY-1999; 99US-0132574P.
 PR 14-MAY-1999; 99US-0132575P.
 PR 14-MAY-1999; 99US-0132576P.
 PR 14-MAY-1999; 99US-0132577P.
 PR 14-MAY-1999; 99US-0132578P.
 PR 14-MAY-1999; 99US-0132579P.
 PR 14-MAY-1999; 99US-0132580P.
 PR 14-MAY-1999; 99US-0132581P.
 PR 14-MAY-1999; 99US-0132582P.
 PR 14-MAY-1999; 99US-0132583P.
 PR 14-MAY-1999; 99US-0132584P.
 PR 14-MAY-1999; 99US-0132585P.
 PR 14-MAY-1999; 99US-0132586P.
 PR 14-MAY-1999; 99US-0132587P.
 PR 14-MAY-1999; 99US-0132588P.
 PR 14-MAY-1999; 99US-0132589P.
 PR 14-MAY-1999; 99US-0132590P.
 PR 14-MAY-1999; 99US-0132591P.
 PR 14-MAY-1999; 99US-0132592P.
 PR 14-MAY-1999; 99US-0132593P.
 PR 14-MAY-1999; 99US-0132594P.
 PR 14-MAY-1999; 99US-0132595P.
 PR 14-MAY-1999; 99US-0132596P.
 PR 14-MAY-1999; 99US-0132597P.
 PR 14-MAY-1999; 99US-0132598P.
 PR 14-MAY-1999; 99US-0132599P.
 PR 14-MAY-1999; 99US-0132600P.
 PR 14-MAY-1999; 99US-0132601P.
 PR 14-MAY-1999; 99US-0132602P.
 PR 14-MAY-1999; 99US-0132603P.
 PR 14-MAY-1999; 99US-0132604P.
 PR 14-MAY-1999; 99US-0132605P.
 PR 14-MAY-1999; 99US-0132606P.
 PR 14-MAY-1999; 99US-0132607P.
 PR 14-MAY-1999; 99US-0132608P.
 PR 14-MAY-1999; 99US-0132609P.
 PR 14-MAY-1999; 99US-0132610P.
 PR 14-MAY-1999; 99US-0132611P.
 PR 14-MAY-1999; 99US-0132612P.
 PR 14-MAY-1999; 99US-0132613P.
 PR 14-MAY-1999; 99US-0132614P.
 PR 14-MAY-1999; 99US-0132615P.
 PR 14-MAY-1999; 99US-0132616P.
 PR 14-MAY-1999; 99US-0132617P.
 PR 14-MAY-1999; 99US-0132618P.
 PR 14-MAY-1999; 99US-0132619P.
 PR 14-MAY-1999; 99US-0132620P.
 PR 14-MAY-1999; 99US-0132621P.
 PR 14-MAY-1999; 99US-0132622P.
 PR 14-MAY-1999; 99US-0132623P.
 PR 14-MAY-1999; 99US-0132624P.
 PR 14-MAY-1999; 99US-0132625P.
 PR 14-MAY-1999; 99US-0132626P.
 PR 14-MAY-1999; 99US-0132627P.
 PR 14-MAY-1999; 99US-0132628P.
 PR 14-MAY-1999; 99US-0132629P.
 PR 14-MAY-1999; 99US-0132630P.
 PR 14-MAY-1999; 99US-0132631P.
 PR 14-MAY-1999; 99US-0132632P.
 PR 14-MAY-1999; 99US-0132633P.
 PR 14-MAY-1999; 99US-0132634P.
 PR 14-MAY-1999; 99US-0132635P.
 PR 14-MAY-1999; 99US-0132636P.
 PR 14-MAY-1999; 99US-0132637P.
 PR 14-MAY-1999; 99US-0132638P.
 PR 14-MAY-1999; 99US-0132639P.
 PR 14-MAY-1999; 99US-0132640P.
 PR 14-MAY-1999; 99US-0132641P.
 PR 14-MAY-1999; 99US-0132642P.
 PR 14-MAY-1999; 99US-0132643P.
 PR 14-MAY-1999; 99US-0132644P.
 PR 14-MAY-1999; 99US-0132645P.
 PR 14-MAY-1999; 99US-0132646P.
 PR 14-MAY-1999; 99US-0132647P.
 PR 14-MAY-1999; 99US-0132648P.
 PR 14-MAY-1999; 99US-0132649P.
 PR 14-MAY-1999; 99US-0132650P.
 PR 14-MAY-1999; 99US-0132651P.
 PR 14-MAY-1999; 99US-0132652P.
 PR 14-MAY-1999; 99US-0132653P.
 PR 14-MAY-1999; 99US-0132654P.
 PR 14-MAY-1999; 99US-0132655P.
 PR 14-MAY-1999; 99US-0132656P.
 PR 14-MAY-1999; 99US-0132657P.
 PR 14-MAY-1999; 99US-0132658P.
 PR 14-MAY-1999; 99US-0132659P.
 PR 14-MAY-1999; 99US-0132660P.
 PR 14-MAY-1999; 99US-0132661P.
 PR 14-MAY-1999; 99US-0132662P.
 PR 14-MAY-1999; 99US-0132663P.
 PR 14-MAY-1999; 99US-0132664P.
 PR 14-MAY-1999; 99US-0132665P.
 PR 14-MAY-1999; 99US-0132666P.
 PR 14-MAY-1999; 99US-0132667P.
 PR 14-MAY-1999; 99US-0132668P.
 PR 14-MAY-1999; 99US-0132669P.
 PR 14-MAY-1999; 99US-0132670P.
 PR 14-MAY-1999; 99US-0132671P.
 PR 14-MAY-1999; 99US-0132672P.
 PR 14-MAY-1999; 99US-0132673P.
 PR 14-MAY-1999; 99US-0132674P.
 PR 14-MAY-1999; 99US-0132675P.
 PR 14-MAY-1999; 99US-0132676P.
 PR 14-MAY-1999; 99US-0132677P.
 PR 14-MAY-1999; 99US-0132678P.
 PR 14-MAY-1999; 99US-0132679P.
 PR 14-MAY-1999; 99US-0132680P.
 PR 14-MAY-1999; 99US-0132681P.
 PR 14-MAY-1999; 99US-0132682P.
 PR 14-MAY-1999; 99US-0132683P.
 PR 14-MAY-1999; 99US-0132684P.
 PR 14-MAY-1999; 99US-0132685P.
 PR 14-MAY-1999; 99US-0132686P.
 PR 14-MAY-1999; 99US-0132687P.
 PR 14-MAY-1999; 99US-0132688P.
 PR 14-MAY-1999; 99US-0132689P.
 PR 14-MAY-1999; 99US-0132690P.
 PR 14-MAY-1999; 99US-0132691P.
 PR 14-MAY-1999; 99US-0132692P.
 PR 14-MAY-1999; 99US-0132693P.
 PR 14-MAY-1999; 99US-0132694P.
 PR 14-MAY-1999; 99US-0132695P.
 PR 14-MAY-1999; 99US-0132696P.
 PR 14-MAY-1999; 99US-0132697P.
 PR 14-MAY-1999; 99US-0132698P.
 PR 14-MAY-1999; 99US-0132699P.
 PR 14-MAY-1999; 99US-0132700P.
 PR 14-MAY-1999; 99US-0132701P.
 PR 14-MAY-1999; 99US-0132702P.
 PR 14-MAY-1999; 99US-0132703P.
 PR 14-MAY-1999; 99US-0132704P.
 PR 14-MAY-1999; 99US-0132705P.
 PR 14-MAY-1999; 99US-0132706P.
 PR 14-MAY-1999; 99US-0132707P.
 PR 14-MAY-1999; 99US-0132708P.
 PR 14-MAY-1999; 99US-0132709P.
 PR 14-MAY-1999; 99US-0132710P.
 PR 14-MAY-1999; 99US-0132711P.
 PR 14-MAY-1999; 99US-0132712P.
 PR 14-MAY-1999; 99US-0132713P.
 PR 14-MAY-1999; 99US-0132714P.
 PR 14-MAY-1999; 99US-0132715P.
 PR 14-MAY-1999; 99US-0132716P.
 PR 14-MAY-1999; 99US-0132717P.
 PR 14-MAY-1999; 99US-0132718P.
 PR 14-MAY-1999; 99US-0132719P.
 PR 14-MAY-1999; 99US-0132720P.
 PR 14-MAY-1999; 99US-0132721P.
 PR 14-MAY-1999; 99US-0132722P.
 PR 14-MAY-1999; 99US-0132723P.
 PR 14-MAY-1999; 99US-0132724P.
 PR 14-MAY-1999; 99US-0132725P.
 PR 14-MAY-1999; 99US-0132726P.
 PR 14-MAY-1999; 99US-0132727P.
 PR 14-MAY-1999; 99US-0132728P.
 PR 14-MAY-1999; 99US-0132729P.
 PR 14-MAY-1999; 99US-0132730P.
 PR 14-MAY-1999; 99US-0132731P.
 PR 14-MAY-1999; 99US-0132732P.
 PR 14-MAY-1999; 99US-0132733P.
 PR 14-MAY-1999; 99US-0132734P.
 PR 14-MAY-1999; 99US-0132735P.
 PR 14-MAY-1999; 99US-0132736P.
 PR 14-MAY-1999; 99US-0132737P.
 PR 14-MAY-1999; 99US-0132738P.
 PR 14-MAY-1999; 99US-0132739P.
 PR 14-MAY-1999; 99US-0132740P.
 PR 14-MAY-1999; 99US-0132741P.
 PR 14-MAY-1999; 99US-0132742P.
 PR 14-MAY-1999; 99US-0132743P.
 PR 14-MAY-1999; 99US-0132744P.
 PR 14-MAY-1999; 99US-0132745P.
 PR 14-MAY-1999; 99US-0132746P.
 PR 14-MAY-1999; 99US-0132747P.
 PR 14-MAY-1999; 99US-0132748P.
 PR 14-MAY-1999; 99US-0132749P.
 PR 14-MAY-1999; 99US-0132750P.
 PR 14-MAY-1999; 99US-0132751P.
 PR 14-MAY-1999; 99US-0132752P.
 PR 14-MAY-1999; 99US-0132753P.
 PR 14-MAY-1999; 99US-0132754P.
 PR 14-MAY-1999; 99US-0132755P.
 PR 14-MAY-1999; 99US-0132756P.
 PR 14-MAY-1999; 99US-0132757P.
 PR 14-MAY-1999; 99US-0132758P.
 PR 14-MAY-1999; 99US-0132759P.
 PR 14-MAY-1999; 99US-0132760P.
 PR 14-MAY-1999; 99US-0132761P.
 PR 14-MAY-1999; 99US-0132762P.
 PR 14-MAY-1999; 99US-0132763P.
 PR 14-MAY-1999; 99US-0132764P.
 PR 14-MAY-1999; 99US-0132765P.
 PR 14-MAY-1999; 99US-0132766P.
 PR 14-MAY-1999; 99US-0132767P.
 PR 14-MAY-1999; 99US-0132768P.
 PR 14-MAY-1999; 99US-0132769P.
 PR 14-MAY-1999; 99US-0132770P.
 PR 14-MAY-1999; 99US-0132771P.
 PR 14-MAY-1999; 99US-0132772P.
 PR 14-MAY-1999; 99US-0132773P.
 PR 14-MAY-1999; 99US-0132774P.
 PR 14-MAY-1999; 99US-0132775P.
 PR 14-MAY-1999; 99US-0132776P.
 PR 14-MAY-1999; 99US-0132777P.
 PR 14-MAY-1999; 99US-0132778P.
 PR 14-MAY-1999; 99US-0132779P.
 PR 14-MAY-1999; 99US-0132780P.
 PR 14-MAY-1999; 99US-0132781P.
 PR 14-MAY-1999; 99US-0132782P.
 PR 14-MAY-1999; 99US-0132783P.
 PR 14-MAY-1999; 99US-0132784P.
 PR 14-MAY-1999; 99US-0132785P.
 PR 14-MAY-1999; 99US-0132786P.
 PR 14-MAY-1999; 99US-0132787P.
 PR 14-MAY-1999; 99US-0132788P.
 PR 14-MAY-1999; 99US-0132789P.
 PR 14-MAY-1999; 99US-0132790P.
 PR 14-MAY-1999; 99US-0132791P.
 PR 14-MAY-1999; 99US-0132792P.
 PR 14-MAY-1999; 99US-0132793P.
 PR 14-MAY-1999; 99US-0132794P.
 PR 14-MAY-1999; 99US-0132795P.
 PR 14-MAY-1999; 99US-0132796P.
 PR 14-MAY-1999; 99US-0132797P.
 PR 14-MAY-1999; 99US-0132798P.
 PR 14-MAY-1999; 99US-0132799P.
 PR 14-MAY-1999; 99US-0132800P.
 PR 14-MAY-1999; 99US-0132801P.
 PR 14-MAY-1999; 99US-0132802P.
 PR 14-MAY-1999; 99US-0132803P.
 PR 14-MAY-1999; 99US-0132804P.
 PR 14-MAY-1999; 99US-0132805P.
 PR 14-MAY-1999; 99US-0132806P.
 PR 14-MAY-1999; 99US-0132807P.
 PR 14-MAY-1999; 99US-0132808P.
 PR 14-MAY-1999; 99US-0132809P.
 PR 14-MAY-1999; 99US-0132810P.
 PR 14-MAY-1999; 99US-0132811P.
 PR 14-MAY-1999; 99US-0132812P.
 PR 14-MAY-1999; 99US-0132813P.
 PR 14-MAY-1999; 99US-0132814P.
 PR 14-MAY-1999; 99US-0132815P.
 PR 14-MAY-1999; 99US-0132816P.
 PR 14-MAY-1999; 99US-0132817P.
 PR 14-MAY-1999; 99US-0132818P.
 PR 14-MAY-1999; 99US-0132819P.
 PR 14-MAY-1999; 99US-0132820P.
 PR 14-MAY-1999; 99US-0132821P.
 PR 14-MAY-1999; 99US-0132822P.
 PR 14-MAY-1999; 99US-0132823P.
 PR 14-MAY-1999; 99US-0132824P.
 PR 14-MAY-1999; 99US-0132825P.
 PR 14-MAY-1999; 99US-0132826P.
 PR 14-MAY-1999; 99US-0132827P.
 PR 14-MAY-1999; 99US-0132828P.
 PR 14-MAY-1999; 99US-0132829P.
 PR 14-MAY-1999; 99US-0132830P.
 PR 14-MAY-1999; 99US-0132831P.
 PR 14-MAY-1999; 99US-0132832P.
 PR 14-MAY-1999; 99US-0132833P.
 PR 14-MAY-1999; 99US-0132834P.
 PR 14-MAY-1999; 99US-0132835P.
 PR 14-MAY-1999; 99US-0132836P.
 PR 14-MAY-1999; 99US-0132837P.
 PR 14-MAY-1999; 99US-0132838P.
 PR 14-MAY-1999; 99US-0132839P.
 PR 14-MAY-1999; 99US-0132840P.
 PR 14-MAY-1999; 99US-0132841P.
 PR 14-MAY-1999; 99US-0132842P.
 PR 14-MAY-1999; 99US-0132843P.
 PR 14-MAY-1999; 99US-0132844P.
 PR 14-MAY-1999; 99US-0132845P.
 PR 14-MAY-1999; 99US-013284

PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147433P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.

PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.3%; Score 266; DB 3; length 288;
Best Local Similarity 28.5%; Pred. No. 3.4e-12;
Matches 82; Conservative 55; Mismatches 119; Indels 32; Gaps 10;

Qy 92 HVDANGKIYLPYLHDWKHPSLELLIQIMIVIFGERPPVFSRFTVSASYPPTATGPPN 151
Db 9 HVTSPGLVSLPYIQNWVYSSNLVDLSLAAAFADPPLYSR---RRQPP--PPSPPT 63
Qy 152 T--SYMEGMPISAYSPPGPNPSPGPGCPYPAGYPATTSQYPSQPPVTTVGFSRD 209
Db 64 VYDSSLRPPSADQSLRPPFPSPYG-----GGVSRVQVQHVHHQQQDDAAAEVFKR 115
Qy 210 GTISE--DTIRASLIASVDKLRWKEEMDGAELNALKRTTEEDLKKGHQKLEENVTR 267
Db 116 NAINKVMVMVHSDIVS--MRRAREAAEELLSQA---GLKRREDELNIG---LKEVV-- 165
Qy 268 LDOEVAEVDKNIELKKKKDELSALEKMEHQSEN---NDIDEVIITAPLYKQILNYA 324
Db 166 --BENETLEQOLQIISNTDILDSWRENOGKTKNLVDLDVNAFECGDTLSKQMLECTA 223
Qy 325 EENAIETIIFYLGEALRRGVIDLDFLKHVRLLSRKQFQLRALMQKAR 372
Db 224 LDLAIEDAIYSLDKSFQDGVVPFDQYLNRVRLLSREQFFHRATGSKVR 271

Search completed: July 12, 2004, 08:35:23
Job time : 58.5664 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2004, 08:29:46 ; Search time 16.5217 Seconds
(without alignments)
1190.526 Million cell updates/sec

Title: US-09-804-690-2

Perfect score: 1996

Sequence: 1 MMSKKYRDLVRQTVNVIA.....FQLRALMOKARKTAGLSLDLY 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/pCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1996	100.0	381	1	US-08-585-758A-2 ✓
2	1996	100.0	381	1	US-08-977-818-2 ✓
3	1996	100.0	381	2	US-08-670-274B-2 ✓
4	1996	100.0	381	2	US-08-786-999-3 ✓
5	1996	100.0	381	3	US-09-146-187-2 ✓
6	1996	100.0	381	4	US-09-216-387-3 ✓
7	1996	100.0	391	1	US-09-886-319A-1
8	1900.5	95.2	380	1	US-08-585-758A-4
9	1900.5	95.2	380	1	US-08-977-818-4
10	1900.5	95.2	380	2	US-08-670-274B-4
11	1900.5	95.2	380	3	US-09-146-187-4
12	1900.5	95.2	390	2	US-08-786-999-1
13	1900.5	95.2	390	4	US-09-216-387-1
14	1900.5	95.2	390	4	US-09-886-319A-2
15	1337.5	67.0	285	4	US-09-886-319A-82
16	1022.5	51.2	237	3	US-08-999-774A-8
17	150	7.5	466	3	US-08-526-136-13
18	146	7.3	148	4	US-09-461-325-453
19	146	7.3	148	4	US-10-012-542-453
20	141.5	7.1	905	2	US-08-574-959A-9
21	141.5	7.1	905	3	US-09-357-014-9
22	141.5	7.1	1135	2	US-08-574-959A-7
23	141.5	7.1	1135	3	US-09-357-014-7
24	125	6.3	591	4	US-09-252-991A-24904
25	122.5	6.1	503	3	US-08-526-136-2
26	122.5	6.1	505	3	US-08-526-136-4
27	121.5	6.1	214	1	US-08-217-327-4

28	121.5	6.1	723	6	5200183-4	Patent No. 5200183
29	120	6.0	1162	2	US-08-728-323A-2	Sequence 2, Appli
30	120	6.0	1162	2	US-09-298-568-2	Sequence 2, Appli
31	120	6.0	1162	4	US-09-410-393-2	Sequence 2, Appli
32	119.5	6.0	816	4	US-09-266-225D-12	Sequence 12, Appli
33	119.5	6.0	8991	4	US-08-714-741-32	Sequence 32, Appli
34	119	6.0	380	2	US-09-026-587-4	Sequence 4, Appli
35	119	6.0	380	2	US-09-227-420-4	Sequence 4, Appli
36	119	6.0	380	4	US-09-387-811-4	Sequence 4, Appli
37	119	6.0	635	4	US-09-417-197-125	Sequence 125, App
38	118.5	5.9	543	4	US-09-535-008-63	Sequence 63, Appli
39	118.5	5.9	559	4	US-10-116-370-2	Sequence 2, Appli
40	118.5	5.9	577	4	US-09-535-008-61	Sequence 61, Appli
41	118.5	5.9	1646	4	US-09-535-008-67	Sequence 67, Appli
42	118.5	5.9	1647	4	US-09-535-008-2	Sequence 2, Appli
43	118.5	5.9	1649	4	US-09-535-008-75	Sequence 75, Appli
44	118.5	5.9	1650	4	US-09-535-008-71	Sequence 71, Appli
45	118.5	5.9	1678	4	US-09-535-008-69	Sequence 69, Appli
46	118.5	5.9	1679	4	US-09-535-008-65	Sequence 65, Appli
47	118.5	5.9	1681	4	US-09-535-008-77	Sequence 77, Appli
48	118.5	5.9	1682	4	US-09-535-008-73	Sequence 73, Appli
49	117.5	5.9	262	3	US-08-946-914-14	Sequence 14, Appli
50	117.5	5.9	262	4	US-09-656-450-14	Sequence 14, Appli
51	117.5	5.9	1184	4	US-09-266-225D-18	Sequence 18, Appli
52	117.5	5.9	1185	3	US-09-041-886-23	Sequence 23, Appli
53	116	5.8	338	1	US-08-218-686-2	Sequence 2, Appli
54	116	5.8	338	3	US-08-460-242-2	Sequence 2, Appli
55	116	5.8	809	5	PCT-US91-01726-3	Sequence 3, Appli
56	115	5.8	652	6	520236-13	Patent No. 520236
57	113	5.7	416	4	US-09-690-454-136	Sequence 136, App
58	113	5.7	779	4	US-10-164-595-56	Sequence 56, Appli
59	113	5.7	843	4	US-10-164-595-54	Sequence 54, Appli
60	112	5.6	272	4	US-09-253-991A-25432	Sequence 25432, A
61	110.5	5.5	558	4	US-09-252-991A-17202	Sequence 17202, A
62	110.5	5.5	902	1	US-08-396-479B-6	Sequence 6, Appli
63	110.5	5.5	902	1	US-08-818-823-6	Sequence 6, Appli
64	110	5.5	1017	4	US-09-600-776-6	Sequence 6, Appli
65	109	5.5	172	4	US-09-800-729-138	Sequence 138, App
66	109	5.5	868	4	US-09-800-729-106	Sequence 106, App
67	109	5.5	921	4	US-09-800-729-199	Sequence 199, App
68	108.5	5.4	531	4	US-09-134-001C-4920	Sequence 4920, Ap
69	108.5	5.4	722	4	US-09-392-714-22	Sequence 22, Appli
70	108.5	5.4	983	4	US-09-311-236-2	Sequence 2, Appli
71	108.5	5.4	1255	2	US-09-080-897-4	Sequence 4, Appli
72	108.5	5.4	1255	3	US-08-899-595-1	Sequence 1, Appli
73	108.5	5.4	1255	3	US-09-323-735-4	Sequence 4, Appli
74	108	5.4	202	6	520236-36	Patent No. 520236
75	108	5.4	203	6	520236-1	Patent No. 520236
76	108	5.4	668	4	US-09-277-431A-2	Sequence 2, Appli
77	107.5	5.4	97	4	US-09-599-360B-118	Sequence 118, App
78	107.5	5.4	2441	1	US-08-194-468-2	Sequence 2, Appli
79	107.5	5.4	2441	3	US-08-961-739-2	Sequence 2, Appli
80	107.5	5.4	2441	4	US-09-514-247A-8	Sequence 8, Appli
81	107.5	5.4	2441	4	US-09-686-316-2	Sequence 2, Appli
82	107	5.4	640	3	US-09-026-343-2	Sequence 2, Appli
83	107	5.4	640	4	US-09-362-871-2	Sequence 2, Appli
84	107	5.4	1461	4	US-09-585-887-9	Sequence 9, Appli
85	107	5.4	1461	4	US-09-289-578-9	Sequence 9, Appli
86	107	5.4	1464	4	US-09-331-347C-21	Sequence 21, Appli
87	106.5	5.3	667	4	US-09-196-270-2	Sequence 2, Appli
88	106.5	5.3	744	6	520236-25	Patent No. 520236
89	105.5	5.3	252	4	US-09-976-594-454	Sequence 454, App
90	105.5	5.3	264	1	US-08-562-311-4	Sequence 4, Appli
91	105	5.3	259	3	US-08-476-509B-29	Sequence 29, Appli
92	105	5.3	1274	4	US-09-095-443-2	Sequence 2, Appli
93	104.5	5.2	529	4	US-08-827-534A-74	Sequence 74, Appli
94	104.5	5.2	529	4	US-09-527-431-74	Sequence 74, Appli
95	104.5	5.2	722	1	US-08-347-718B-1	Sequence 1, Appli
96	104.5	5.2	722	1	US-08-445-050-3	Sequence 3, Appli
97	104.5	5.2	722	1	US-08-445-050-7	Sequence 7, Appli
98	104.5	5.2	722	2	US-08-482-262-1	Sequence 1, Appli
99	104.5	5.2	722	2	US-08-204-691-3	Sequence 3, Appli
100	104.5	5.2	722	2	US-08-204-691-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-08-585-758A-2
; Sequence 2, Application US/08585758A
; Patent No. 5679523
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N.
; TITLE OF INVENTION: METHOD FOR CONCURRENT DISRUPTION OF
; TITLE OF INVENTION: EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHER, HOEBACH, TEST, ALBERTSON & HERBERT
; STREET: FOUR EMBARCADERO CENTER, SUITE 3400
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 12-JAN-1996
; APPLICATION NUMBER: US/08/585,758A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A62783/BI
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-585-758A-2

Query Match 100.0%; Score 1996; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 4.6e-158;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MMSKYKRDLTVRQTVNVIAMYKDKLPVDSYVFNDSGSSRELNLGTIPVRYEGNIYNI	60
Db	1	MMSKYKRDLTVRQTVNVIAMYKDKLPVDSYVFNDSGSSRELNLGTIPVRYEGNIYNI	60
Qy	61	PICLWLDDTYPNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELLELIQI	120
Db	61	PICLWLDDTYPNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELLELIQI	120
Qy	121	MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGPGISAYPSGYPNPSGPGCP	180
Db	121	MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGPGISAYPSGYPNPSGPGCP	180
Qy	181	YPPAGPYPATSSQPSQPPVTTVGPREDGTISEDTIRASLISAVSKLRWKEEMDGA	240
Db	181	YPPAGPYPATSSQPSQPPVTTVGPREDGTISEDTIRASLISAVSKLRWKEEMDGA	240
Qy	241	QAEINLAKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELKKKDEELSSALEKMNQS	300
Db	241	QAEINLAKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELKKKDEELSSALEKMNQS	300
Qy	301	ENNDIDEVITFAPLYKQILNLYAEENAIETIIFVLGEALRGVIDLDVFLKRVLLSRK	360
Db	301	ENNDIDEVITFAPLYKQILNLYAEENAIETIIFVLGEALRGVIDLDVFLKRVLLSRK	360

Qy 361 QFQLRALMOKARKTAGLSLDLY 381
Db 361 QFQLRALMOKARKTAGLSLDLY 381

RESULT 2

US-08-977-818-2
; Sequence 2, Application US/08977818
; Patent No. 5807995
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N.
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 25-NOV-1997
; APPLICATION NUMBER: US/08/977,818
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-977-818-2

Query Match 100.0%; Score 1996; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 4.6e-158;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MMSKYKRDLTVRQTVNVIAMYKDKLPVDSYVFNDSGSSRELNLGTIPVRYEGNIYNI	60
Db	1	MMSKYKRDLTVRQTVNVIAMYKDKLPVDSYVFNDSGSSRELNLGTIPVRYEGNIYNI	60
Qy	61	PICLWLDDTYPNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELLELIQI	120
Db	61	PICLWLDDTYPNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELLELIQI	120
Qy	121	MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGPGISAYPSGYPNPSGPGCP	180
Db	121	MIVIFGEEPPVFSRPTVSASYPPTATGPPNTSYMPGPGISAYPSGYPNPSGPGCP	180
Qy	181	YPPAGPYPATSSQPSQPPVTTVGPREDGTISEDTIRASLISAVSKLRWKEEMDGA	240
Db	181	YPPAGPYPATSSQPSQPPVTTVGPREDGTISEDTIRASLISAVSKLRWKEEMDGA	240
Qy	241	QAEINLAKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELKKKDEELSSALEKMNQS	300
Db	241	QAEINLAKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELKKKDEELSSALEKMNQS	300

QY 301 ENNDIDEVIPTAPLYKQILNLYAEENAIETIFVLGEALRGVIDLDVFLKHVRLLSRK 360
Db 301 ENNDIDEVIPTAPLYKQILNLYAEENAIETIFVLGEALRGVIDLDVFLKHVRLLSRK 360
QY 361 QFQRLALMQKARKTAGLSLDLY 381
Db 361 QFQRLALMQKARKTAGLSLDLY 381

RESULT 3
US-08-670-274B-2
; Sequence 2, Application US/08670274B
; Patent No. 5891668
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,274B
; FILING DATE: June 13, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-670-274B-2

Query Match 100.0%; Score 1996; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 4.6e-158;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMSKYKYRDLTVRQTVNVNVIAMYKDKLPVLDYSYVFDGSSRELNVLTGTIPVYRGNIYNI 60
Db 1 MMSKYKYRDLTVRQTVNVNVIAMYKDKLPVLDYSYVFDGSSRELNVLTGTIPVYRGNIYNI 60
QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELELIQI 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELELIQI 120
QY 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSPGPGCP 180
Db 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSPGPGCP 180
QY 181 YPPAGPYPATTSQYPSQPPVTVGSPRDGHTISETTIRASLISAVSOKLRWRKMEEMDGA 240
Db 181 YPPAGPYPATTSQYPSQPPVTVGSPRDGHTISETTIRASLISAVSOKLRWRKMEEMDGA 240
QY 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVDNKNIELLKKKDEELSSALEKMEENQS 300

Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVDNKNIELLKKKDEELSSALEKMEENQS 300
QY 301 ENNDIDEVIPTAPLYKQILNLYAEENAIETIFVLGEALRGVIDLDVFLKHVRLLSRK 360
Db 301 ENNDIDEVIPTAPLYKQILNLYAEENAIETIFVLGEALRGVIDLDVFLKHVRLLSRK 360
QY 361 QFQRLALMQKARKTAGLSLDLY 381
Db 361 QFQRLALMQKARKTAGLSLDLY 381

RESULT 4
US-08-786-999-3
; Sequence 3, Application US/08786999
; Patent No. 5892016
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; TITLE OF INVENTION: SUPPRESSOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: Filed Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1330330
US-08-786-999-3

Query Match 100.0%; Score 1996; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 4.6e-158;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMSKYKYRDLTVRQTVNVNVIAMYKDKLPVLDYSYVFDGSSRELNVLTGTIPVYRGNIYNI 60
Db 1 MMSKYKYRDLTVRQTVNVNVIAMYKDKLPVLDYSYVFDGSSRELNVLTGTIPVYRGNIYNI 60
QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELELIQI 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTKGHDVANGKIYLPYLHDWKHPRSELELIQI 120
QY 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSPGPGCP 180
Db 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSPGPGCP 180

Qy	181	YPPAGPYPATSSQVPSOPPVTTVGSPRDTISEDTIRASLISAVSDKLRWRMKEMDGA	240
Db	181	YPPAGPYPATSSQVPSOPPVTTVGSPRDTISEDTIRASLISAVSDKLRWRMKEMDGA	240
Qy	241	QAEINALKRTBEDLKKGHOKLEEMVTRLDQVAEVDKIELLKKDEELSSALEKXWQS	300
Db	241	QAEINALKRTBEDLKKGHOKLEEMVTRLDQVAEVDKIELLKKDEELSSALEKXWQS	300
Qy	301	ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRGVTDLDVFLKRVLLSRK	360
Db	301	ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRGVTDLDVFLKRVLLSRK	360
Qy	361	QFQRLMQKARKTAGLSLDLY	381
Db	361	QFQRLMQKARKTAGLSLDLY	381

RESULT 5
 US-09-146-187-2
 ; Sequence 2, Application US/09146:187
 ; Patent No. 6248523
 ; GENERAL INFORMATION:
 ; APPLICANT: LI, Limin
 ; APPLICANT: COHEN, Stanley N
 ; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
 ; TITLE OF INVENTION: THEIR USES
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FISH AND RICHARDSON, P.C.
 ; STREET: 2200 SAND HILL ROAD
 ; CITY: MENLO PARK
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/146,187
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/670,274
 ; FILING DATE: June 13, 1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: SHERWOOD, Pamela J.
 ; REGISTRATION NUMBER: 36,677
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-781-1989
 ; TELEFAX: 415-398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 381 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-09-146-187-2

Qy	121	MIVIFGEEPPVFSRPTVSASVPPYTATGPPNTSYMPGMPSGISAYSPSGYPPNPSPGYGCP	180
Db	121	MIVIFGEEPPVFSRPTVSASVPPYTATGPPNTSYMPGMPSGISAYSPSGYPPNPSPGYGCP	180
Qy	181	YPPAGVPATTSSQVPSOPPVTTVGPSRDGITSEDITIRASLISAYSDDLKRWKKEEMDCA	240
Db	181	YPPAGVPATTSSQVPSOPPVTTVGPSRDGITSEDITIRASLISAYSDDLKRWKKEEMDCA	240
Qy	241	QAEINALKRTBEDLKGHQKLEEMVTRLDQDVAEVDKNIELLKKDEELSSALEKWNOS	300
Db	241	QAEINALKRTBEDLKGHQKLEEMVTRLDQDVAEVDKNIELLKKDEELSSALEKWNOS	300
Qy	301	ENNDIDEVIIITAPLYKOILNLYAEENAIEDTIFVLGEALRGVTDLOVFLKHVRLSPK	360
Db	301	ENNDIDEVIIITAPLYKOILNLYAEENAIEDTIFVLGEALRGVTDLOVFLKHVRLSPK	360
Qy	361	QFOLRALMQARKTAGTSDLY	381
Db	361	QFOLRALMQARKTAGTSDLY	381

RESULT 6
US-09-216-387-3
; Sequence 3, Application US/09216387
; Patent No. 6472508
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; SUPPRESSOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/216.387
; FILING DATE: 18-Dec-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: J330330
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-216-387-3

QY 1 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYSYVNDGSSRELNLGTGTPVRYRGNINYI 60
Db 1 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYSYVNDGSSRELNLGTGTPVRYRGNINYI 60
QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDWKHPRSELLELIQI 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDWKHPRSELLELIQI 120
QY 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180
Db 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180
QY 181 YPPAGPYPATSSQYPSQPPVTVGSPRDGTISEDTRASLISAVSOKLRWRMKEEMDGA 240
Db 181 YPPAGPYPATSSQYPSQPPVTVGSPRDGTISEDTRASLISAVSOKLRWRMKEEMDGA 240
QY 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDBELSSALEKMNQNS 300
Db 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDBELSSALEKMNQNS 300
QY 301 ENNDIDEVIIPTAPLYKQILNLYAENAIETIFVLGEALRRGVLDLDVFLKHVRLLSRK 360
Db 301 ENNDIDEVIIPTAPLYKQILNLYAENAIETIFVLGEALRRGVLDLDVFLKHVRLLSRK 360
QY 361 QFQRLALMQKARKTAGLSLDLY 381
Db 361 QFQRLALMQKARKTAGLSLDLY 381
RESULT 7
US-09-886-319A-1
; Sequence 1, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jörn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886.319A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-886-319A-1
Query Match 100.0%; Score 1996; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.7e-158;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYSYVNDGSSRELNLGTGTPVRYRGNINYI 60
Db 11 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYSYVNDGSSRELNLGTGTPVRYRGNINYI 70
QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDWKHPRSELLELIQI 120
Db 71 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDWKHPRSELLELIQI 130
QY 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180

Db 131 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 190
QY 181 YPPAGPYPATSSQYPSQPPVTVGSPRDGTISEDTRASLISAVSOKLRWRMKEEMDGA 240
Db 191 YPPAGPYPATSSQYPSQPPVTVGSPRDGTISEDTRASLISAVSOKLRWRMKEEMDGA 250
QY 241 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDBELSSALEKMNQNS 300
Db 251 QAEINALKRTEEDLKKGHQKLEEMVTRLDQVAVDKNIELLKKKDBELSSALEKMNQNS 310
QY 301 ENNDIDEVIIPTAPLYKQILNLYAENAIETIFVLGEALRRGVLDLDVFLKHVRLLSRK 360
Db 311 ENNDIDEVIIPTAPLYKQILNLYAENAIETIFVLGEALRRGVLDLDVFLKHVRLLSRK 370
QY 361 QFQRLALMQKARKTAGLSLDLY 381
Db 371 QFQRLALMQKARKTAGLSLDLY 391
RESULT 8
US-08-585-758A-4
; Sequence 4, Application US/08585758A
; Patent No. 5679523
; GENERAL INFORMATION:
; APPLICANT: Ili, Limin
; APPLICANT: COHEN, Stanley N.
; TITLE OF INVENTION: METHOD FOR CONCURRENT DISRUPTION OF
; TITLE OF INVENTION: EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: FOUR EMBARCADERO CENTER, SUITE 3400
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/585,758A
; APPLICATION NUMBER: US/08/585,758A
; FILING DATE: 12-JAN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A62783/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-585-758A-4
Query Match 95.2%; Score 1900.5; DB 1; Length 380;
Best Local Similarity 94.5%; Pred. No. 4e-150;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYSYVNDGSSRELNLGTGTPVRYRGNINYI 60
Db 1 MMSKYKYRDLTVRQTVNVIAMKYDKLPVLDYSYVNDGSSRELNLGTGTPVRYRGNINYI 60
QY 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDWKHPRSELLELIQI 120
Db 61 PICLWLLDTPYNPPICFVKPTSSMTIKTGHVDANGKIYLPYLHDWKHPRSELLELIQI 120
QY 121 MIVIFGEPPVFSRPTVSASYPPTATGPPNTSYMPGMPGSGISAYPSGYPNPSGYPGCP 180

Db 121 MIVVFGDEPPVSRP-ISASYPYQATGPNTSYMEFGPGGISPYSGYPNPSGYPGCP 179
QY 181 YYPAGYPATTSSQSPQPPVTTVGSRDGTISEDITIRASLISAVSDKLRWRKEMDGA 240
Db 180 YPPGGYPATTSSQSPQPPVTTVGSRDGTISEDITIRASLISAVSDKLRWRKEMDRA 239
QY 241 QAEINALKTEEDLKKGHQKLEEMVTRLDQEAQVAEVDKNIELKKDBEELSALAKMENOS 300
Db 240 QAEINALKTEEDLKKGHQKLEEMVTRLDQEAQVAEVDKNIELKKDBEELSALAKMENOS 299
QY 301 ENNDIDEVIPTAPLYKQILNLYAEENAIETDIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 300 ENNDIDEVIPTAPLYKQILNLYAEENAIETDIFYLGEALRRGVLDLDFLKHVRLLSRK 359
QY 361 QFQLRALMOKARKTAGLSLDLY 381
Db 360 QFQLRALMOKARKTAGLSLDLY 380

RESULT 9

US-08-977-818-4
; Sequence 4, Application US/08977818
; Patent No. 5807995
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 25-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-977-818-4

Query Match 95.2%; Score 1900.5; DB 1; Length 380;
Best Local Similarity 94.5%; Pred. No. 4e-150; 9; Indels 1; Gaps 1;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MMSKYRDLTVRQTVNVIAMKYKLPVLDSDYVFNDSGSSRELNVNLGTIPVYRGNIYNI 60
Db 1 MVSXYRDLTVRETWNVITLYKDLKPVLDSDYVFNDSGSSRELNVNLGTIPVYRGNTYNI 60
QY 61 PCLWLLDTYPPNPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDKWKPRLLELIQI 120

Db 61 PCLWLLDTYPPNPICFVKPTSSMTIKTKGHVDANGKIYLPYLHDKWKPRLLELIQI 120
QY 121 MIVVFGDEPPVSRP-ISASYPYQATGPNTSYMEFGPGGISPYSGYPNPSGYPGCP 180
Db 121 MIVVFGDEPPVSRP-ISASYPYQATGPNTSYMEFGPGGISPYSGYPNPSGYPGCP 179
QY 181 YYPAGYPATTSSQSPQPPVTTVGSRDGTISEDITIRASLISAVSDKLRWRKEMDGA 240
Db 180 YPPGGYPATTSSQSPQPPVTTVGSRDGTISEDITIRASLISAVSDKLRWRKEMDRA 239
QY 241 QAEINALKTEEDLKKGHQKLEEMVTRLDQEAQVAEVDKNIELKKDBEELSALAKMENOS 300
Db 240 QAEINALKTEEDLKKGHQKLEEMVTRLDQEAQVAEVDKNIELKKDBEELSALAKMENOS 299
QY 301 ENNDIDEVIPTAPLYKQILNLYAEENAIETDIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 300 ENNDIDEVIPTAPLYKQILNLYAEENAIETDIFYLGEALRRGVLDLDFLKHVRLLSRK 359
QY 361 QFQLRALMOKARKTAGLSLDLY 381
Db 360 QFQLRALMOKARKTAGLSLDLY 380

RESULT 10

US-08-670-274B-4
; Sequence 4, Application US/08670274B
; Patent No. 5891668
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,274B
; FILING DATE: June 13, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-670-274B-4

Query Match 95.2%; Score 1900.5; DB 2; Length 380;
Best Local Similarity 94.5%; Pred. No. 4e-150; 9; Indels 1; Gaps 1;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;
QY 1 MMSKYRDLTVRQTVNVIAMKYKLPVLDSDYVFNDSGSSRELNVNLGTIPVYRGNIYNI 60
Db 1 MVSXYRDLTVRETWNVITLYKDLKPVLDSDYVFNDSGSSRELNVNLGTIPVYRGNTYNI 60


```
; CLONE: 609476
US-08-786-999-1

Query Match          95.2%; Score 1900.5; DB 2; Length 390;
Best Local Similarity 94.5%; Pred. No. 4.2e-150;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDYSYVFDGSSRELNLNLTGTIPVYRGNIYNI 60
Db 11 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDYSYVFDGSSRELNLNLTGTIPVYRGNIYNI 70

QY 61 PICMLLDDTYPNPPICFVKPTSSMTIKTKGKHDVANGKIYLPYLHDKHPRSELLELIQI 120
Db 71 PICMLLDDTYPNPPICFVKPTSSMTIKTKGKHDVANGKIYLPYLHDKHPRSELLELIQI 130

QY 121 MIVFGEPPVFRSPRTVSASYPYATGPPNTSYMPGMPGSGISAYPSGYPPNPSGYPGCP 180
Db 131 MIVFGEPPVFRSPRTVSASYPYATGPPNTSYMPGMPGSGISAYPSGYPPNPSGYPGCP 189

QY 181 YPAGYPATTSSQYPSQPPVTTVGPSRDGTISEDITIRASLISAVSDKLRWRKKEMDGA 240
Db 190 YPAGYPATTSSQYPSQPPVTTVGPSRDGTISEDITIRASLISAVSDKLRWRKKEMDRA 249

QY 241 QAEINAKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNIELLKKDEELSSALEKMNQNS 300
Db 250 QAEINAKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNIELLKKDEELSSALEKMNQNS 309

QY 301 ENNDIDEVIPTAPLYKQILNLNLAENNAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 310 ENNDIDEVIPTAPLYKQILNLNLAENNAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 369

QY 361 QFQRLALMOKARKTAGLSLDLY 381
Db 370 QFQRLALMOKARKTAGLSLDLY 390

RESULT 13
US-09-216-387-1
; Sequence 1, Application US/09216387
; Patent No. 6472508
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; SUPPRESSOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/216,387
; FILING DATE: 18-Dec-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Colnot01
; CLONE: 609476
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-216-387-1

Query Match          95.2%; Score 1900.5; DB 4; Length 390;
Best Local Similarity 94.5%; Pred. No. 4.2e-150;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDYSYVFDGSSRELNLNLTGTIPVYRGNIYNI 60
Db 11 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDYSYVFDGSSRELNLNLTGTIPVYRGNIYNI 70

QY 61 PICMLLDDTYPNPPICFVKPTSSMTIKTKGKHDVANGKIYLPYLHDKHPRSELLELIQI 120
Db 71 PICMLLDDTYPNPPICFVKPTSSMTIKTKGKHDVANGKIYLPYLHDKHPRSELLELIQI 130

QY 121 MIVFGEPPVFRSPRTVSASYPYATGPPNTSYMPGMPGSGISAYPSGYPPNPSGYPGCP 180
Db 131 MIVFGEPPVFRSPRTVSASYPYATGPPNTSYMPGMPGSGISAYPSGYPPNPSGYPGCP 189

QY 181 YPAGYPATTSSQYPSQPPVTTVGPSRDGTISEDITIRASLISAVSDKLRWRKKEMDGA 240
Db 190 YPAGYPATTSSQYPSQPPVTTVGPSRDGTISEDITIRASLISAVSDKLRWRKKEMDRA 249

QY 241 QAEINAKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNIELLKKDEELSSALEKMNQNS 300
Db 250 QAEINAKRTEEDLKKGHQKLEEMVTRLDQEAEDVDKNIELLKKDEELSSALEKMNQNS 309

QY 301 ENNDIDEVIPTAPLYKQILNLNLAENNAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 360
Db 310 ENNDIDEVIPTAPLYKQILNLNLAENNAIEDTIFYLGEALRRGVLDLDFLKHVRLLSRK 369

QY 361 QFQRLALMOKARKTAGLSLDLY 381
Db 370 QFQRLALMOKARKTAGLSLDLY 390

RESULT 14
US-09-886-319A-2
; Sequence 2, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; the Diagnosis or Treatment of Skin Disorders and Wound
; Healing and for the Identification of Pharmacologically
; Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-319A-2
```

```
Query Match      95.2%; Score 1900.5; DB 4; Length 390;
Best Local Similarity 94.5%; Pred. No. 4.2e-150;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MMSKYKYRDLTVRQVNVNIAMVKDLKPVLDVSFVNDGSSRELNVLTGTIPVYRGNIYNI 60
DB 11 MYSKYKYRDLTVRETVNVITLYKDLKPVLDVSFVNDGSSRELNVLTGTIPVYRGNTYNI 70

QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHDVANGKIYLYPYLHDWKHPRSELLELIQI 120
DB 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHDVANGKIYLYPYLHEWKHPQSDLLGLIQV 130

QY 121 MIVIFGEPPVPSRTVSASYPYATGPPNTSYMPGMPGSGISAYPSGYPNPSPGPGCP 180
DB 131 MIVIFGEPPVPSR- ISASYPYQATGPPNTSYMPGMPGSGISAYPSGYPNPSPGPGCP 189

QY 181 YPPAGPYPATTSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKLRWKMKEMDGA 240
DB 190 YPPGPGYPATTSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKLRWKMKEMDRA 249

QY 241 QAEINALKRTEBDLKKGHQKLEEMVTRLDQVEAEVDKNIELKKXDEELSSALEKXENQS 300
DB 250 QAEINALKRTEBDLKKGHQKLEEMVTRLDQVEAEVDKNIELKKXDEELSSALEKXENQS 309

QY 301 ENNDIDEVIPTAPLYKQILNLYAEENAELEDTFYLGALRRGVLDLDFLKHVRLLSRK 360
DB 310 ENNDIDEVIPTAPLYKQILNLYAEENAELEDTFYLGALRRGVLDLDFLKHVRLLSRK 369

QY 361 QFQLRALMQARKTAGLSLDLY 381
DB 370 QFQLRALMQARKTAGLSLDLY 390

RESULT 15
US-09-886-319A-82
; Sequence 82, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-319A-82

Query Match      67.0%; Score 1337.5; DB 4; Length 285;
Best Local Similarity 95.2%; Pred. No. 1.9e-103;
Matches 259; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

QY 110 PRSELLELIQIMIVIFGEPPVPSRTVSASYPYATGPPNTSYMPGMPGSGISAYPSGY 169
DB 15 PQSDLLGLIQIMIVIFGEPPVPSR- ISASYPYQATGPPNTSYMPGMPGSGISAYPSGY 73

QY 170 PNPSPGPGCPYPAGPYPATTSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKL 229
DB 74 PNPSPGPGCPYPAGPYPATTSQYPSQPPVTTVGPSSRDGTISEDTIRASLISAVSDKL 133

QY 230 RWRMKEEMDGAQAEINALKRTEBDLKKGHQKLEEMVTRLDQVEAEVDKNIELKKXDEEL 289
DB 134 RWRMKEEMDRAQAEINALKRTEBDLKKGHQKLEEMVTRLDQVEAEVDKNIELKKXDEEL 193

QY 290 SSALEKXENQSENNDIDEVIPTAPLYKQILNLYAEENAELEDTFYLGALRRGVLDLDF 349
DB 194 SSALEKXENQSENNDIDEVIPTAPLYKQILNLYAEENAELEDTFYLGALRRGVLDLDF 253

QY 350 FLKHVRLLSRKQFQLRALMQARKTAGLSLDLY 381
DB 254 FLKHVRLLSRKQFQLRALMQARKTAGLSLDLY 285

RESULT 16
US-08-999-774A-8
; Sequence 8, Application US/08999774A
; Patent No. 6274312
; GENERAL INFORMATION:
; APPLICANT: Gish, Kurt C.
; APPLICANT: Seghezzi, Wolfgang
; APPLICANT: Shanahan, Frances
; APPLICANT: Lees, Emma M.
; APPLICANT: McClanahan, Terrill K.
; TITLE OF INVENTION: Intracellular Regulatory Molecules;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,774A
; FILING DATE: 10-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,818
; FILING DATE: 11-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 852-9196
; TELEFAX: (650) 496-1200
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-999-774A-8

Query Match      51.2%; Score 1022.5; DB 3; Length 237;
Best Local Similarity 88.3%; Pred. No. 2.4e-77;
Matches 182; Conservative 12; Mismatches 11; Indels 1; Gaps 1;

QY 1 MMSKYKYRDLTVRQVNVNIAMVKDLKPVLDVSFVNDGSSRELNVLTGTIPVYRGNIYNI 60
DB 11 MYSKYKYRDLTVRETVNVITLYKDLKPVLDVSFVNDGSSRELNVLTGTIPVYRGNTYNI 70

QY 61 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHDVANGKIYLYPYLHDWKHPRSELLELIQI 120
DB 71 PICLWLLDTPYNNPPICFVKPTSSMTIKTGKHDVANGKIYLYPYLHEWKHPQSDLLGLIQV 130
```

QY 121 MIVIFGEPPVFRSRTVSASYPYATGTPNTSYMPGMSGISAYPSGYPNPSGYGPGCP 180
Db 131 MIVVFGDEPPVFSRP-ISASYPYQATGTPNTSYMPGMSGISAYPSGYPNPSGYGPGCP 189
QY 181 YPPAGPYPATSSQYPSQPPVTVGP 206
Db 190 YPPGGYPATSSQYPSQPPVPLVP 215
RESULT 17
US-08-526-136-13
; Sequence 13, Application US/08526136
; Patent No. 6107089
; GENERAL INFORMATION:
; APPLICANT: Towle, Christine A. et al.
; TITLE OF INVENTION: ANNEXIN XI
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,136
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,036
; FILING DATE:
; APPLICATION NUMBER: 07/837,775
; FILING DATE: February 13, 1992
; APPLICATION NUMBER: 07/764,465
; FILING DATE: September 23, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/099001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-08-526-136-13
Query Match 7.5%; Score 150; DB 3; Length 466;
Best Local Similarity 21.3%; Pred. No. 0.00026;
Matches 83; Conservative 36; Mismatches 92; Indels 178; Gaps 16;
QY 140 SYPPYATG-PPNTSYMPG-----MPSGISAYPSGYG----- 170
Db 2 SYFGYPTGYPPFPYPPAGQSSFFPSGGYPPGPGGAYPQVPSSGYGAGYGP 61
QY 171 -----FNPSSYPCPPYAGP-YPATSSQ-----YPSQPPVTVG----- 205
Db 62 APGGYAPGYPGAPQPGGAPSPYGPVPGGFGVPPGAGFSGYP-QPPSQSYGGGPAQV 120
QY 206 -----PSRDGTISED----- 216
Db 121 PLPGGFGGMPGQYPGGQTYPSQPATVTVGTGTRPAANFATDAIRKAMKGF 180
QY 217 -----IRASLISAVSDKLRWRMKEEMDGAQAEALNA---LKRTEED 253

Db 181 TDEQAIVDVVANSNDQRQKIAFKTSYKGLIKDKLSLSGNMBELIILALFMPPTYYD 240
QY 254 ---LKGHQ-----KLEFMYRLDDQEAHV-----DNIELLKKKDEELSSALEKM 296
Db 241 AWSLRKAMQAGTQERVLIEILTRTNQEIREFVRCVQSEFGSDLEKDIRSDTSGHFERL 300
QY 297 -----ENQSENNDIDEVIIPTAPLYKQILNLYAENAIETIFYLGEALRRVID 346
Db 301 LVSMCQGNRDENQSIHQ-----AQEDA---QRLYQAGE-----CRLG 336
QY 347 LDVFLKHVLLSRKQFQLRALMOKARKTA 375
Db 337 TDESCFNMLLATRSFFQLRATMEAYSMA 365
RESULT 18
US-09-461-325-453
; Sequence 453, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-325-453
Query Match 7.3%; Score 146; DB 4; Length 148;
Best Local Similarity 27.9%; Pred. No. 0.00011;
Matches 38; Conservative 15; Mismatches 33; Indels 50; Gaps 6;
QY 135 PTVSASYPYATGTPNTSYMPGMSGISAYPSGYGPNP---SGYPCG----PYPPAGPY 187
Db 30 PPINPPFPFGPCPPP-----PGAPHGNPAFFPGGPHVPVQPGYGCQPLGYPYPPYP 83
QY 188 PATTSSQYPSQPPVTVGSRDGTISEDITIRASLISAVSDKLRWRMKEEMDGAQAEALNA 247
Db 84 PA-----PGIPPNPLADGMVGP-----AVVD----- 106
QY 248 KRTEEDLKKGHOKLEE 263
Db 107 KMQQKMKKAKKMKHK 122
RESULT 19
US-10-012-542-453
; Sequence 453, Application US/10012542
; Patent No. 6627741
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1

```
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-012-542-453

Query Match          7.3%; Score 146; DB 4; Length 148;
Best Local Similarity 27.9%; Pred. No. 0.00011;
Matches 38; Conservative 15; Mismatches 33; Indels 50; Gaps 6;

QY 135 PTVSASYPPTATGPPNTSYMPGMPGSGAYSQVPPNP---SGYPGC---PYPPAGPY 187
DB 30 PPINPPFPFGCPPP-----FGAPHGNFAFPFGPPHVPQGVYGCQPLGFPYPPYP 83
QY 188 PATTSQSPSPQPVTVVGSRSRGTSTEDTIRASLISAVSKLWRMKEMDGAQELNAL 247
DB 84 PA-----PGIPFVNPLAPGMVGP-----AVIVD----- 106
QY 248 KTEEDLKKGHOKLEE 263
DB 107 KQVQKMKKKAHKKMKH 122

RESULT 20
US-08-574-959A-9
; Sequence 9, Application US/08574959A
; Patent No. 5962224
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joungh, Ratna K. Vadlamudi
; APPLICANT: and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/574,959A
; FILING DATE: 19-DEC-95
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
```

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-574-959A-9

Query Match          7.1%; Score 141.5; DB 2; Length 905;
Best Local Similarity 26.0%; Pred. No. 0.0034;
Matches 50; Conservative 32; Mismatches 83; Indels 27; Gaps 7;

QY 122 IVIFGEE-PPVFSRPTVSASYPPTATGPPNTSYMPGMPG-----GISAYPSGYPPNPSG 175
DB 554 WVIVPEGLPPLPPPPSGATPPPIAPTGPPTAS--PPVPAKEEPELPAAPGLPPPP-- 609
QY 176 YPGCPYPAGPYPATTSQSPSPQPVTVVGSRSRGTSTEDTIRASLISAVSKLWRMKH 235
DB 610 -----PPPPVPGPVXLPPLPVPEGTGGGGGPALEEDLTVININS--SDE-----EE 656
QY 236 EMDGAQELNALKRTEDLKKGHOKLEEVMVTRLDQVAVDVKNIELLKKKDELSALEK 295
DB 657 EEEGEEEE---EEEEEEEEEEEEEEEDFEEDEEYFEEEEEEEEEEFEFE 712
QY 296 MENQSENNDIDE 307
DB 713 EEEGEEEEEEEE 724

RESULT 21
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joungh, Ratna K. Vadlamudi
; APPLICANT: and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,014
; FILING DATE: 19-Jul-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/574,959
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
; US-09-357-014-9
```


This Page Blank (uspto)